Niveau(x) d'enseignement: 0 Numéro EDUQ : 8570349
Centre d'intérêt: 841

Bertrand, Richard.

Comparaison de diverses méthodes d'analyse factorielle / Richard Bertrand, Normand Dufour. -- [Sainte-Foy] : INRS-Education, 1984.
v, 141 p. : graph. ; 28 cm.

Code: R-170. Bibliogr.: p. [129]-133.

Descripteur(s): *Méthode d'analyse *Analyse de variables *Evaluation d'une méthode *Etude comparative *Statistique

Identificateur(s) de sujet: *Analyse factorielle

L'analyse factorielle peut être décrite comme un ensemble de méthodes servant soit à classer les variables en tenant compte de leurs associations (ou corrélations), soit à classer les sujets en tenant compte des relations entre leurs scores à diverses variables. L'analyse factorielle constitue aujourd'hui encore une des méthodes d'analyse les plus utilisées. Mais sous le vocable "analyse factorielle" se cache non pas une mais bien plusieurs méthodes d'analyse. Le problème qui se pose est celui de l'utilisation de l'analyse factorielle. Plus précisément, il s'agit de déterminer la méthode d'analyse factorielle qui convienne à un type particulier de données. L'étude vise à faciliter le choix d'une méthode.

Avant d'aborder cette problématique, les auteurs ont jugé utile de faire une présentation sommaire, visuelle et intuitive, de l'analyse factorielle. Cette présentation constitue la première partie du texte. La deuxième partie de l'étude aborde la question du choix d'une méthode, d'un point de vue empirique, en présentant une étude comparative de diverses méthodes d'analyse factorielle.

Au terme de leur étude, les auteurs posent un regard quelque peu perplexe sur les résultats obtenus. Le bilan n'est pas vraiment négatif, mais ils n'ont pas trouvé la meilleure solution pour chacun des types de matrices de corrélations étudiés tel qu'ils l'escomptaient au début de leur recherche. Ils sont d'avis néanmoins que tout bon "utilisateur modéré" devrait pouvoir discriminer entre les méthodes acceptables et celles qui ne le sont pas, compte tenu des données dont il dispose. La bonne question à se poser pour le choix d'une méthode serait: "Est-ce que cette méthode paraît acceptable?" plutôt que: "Quelle est la meilleure solution?".

Type de document: r
Financement: Bénéficiaire - INRS-Education
Programme(s) de subvention - FCAC

Éducation

EDUQ 8570349

COMPARAISON DE DIVERSES MÉTHODES D'ANALYSE FACTORIELLE

Richard Bertrand
Normand Dufour

R-170

1984

INRS-Éducation
2383 chemin Ste-Foy
Ste-Foy
G1V 1T1

REMERCIEMENTS

Nous désirons adresser nos remerciements aux organismes F.C.A.C. (Québec) et C.R.S.N.G. (Ottawa) qui ont bien voulu subventionner nos recherches.

Par ailleurs nous remercions sincèrement M. Djavid Ajar, professeur à l'Université de Montréal, Jean-Jacques Bernier, professeur à l'Université Laval, et Horacio Rimoldi, directeur du C.I.I.P.M. (Buenos Aires, Argentine) qui ont accepté, à titre gracieux, d'agir comme juges.

Enfin nos remerciements vont au personnel de l'INRS-Éducation qui a collaboré à la recherche, et en particulier au professeur Yves Bégin qui a participé à cette étude depuis ses tous débuts ainsi qu'à Madame Simone Roy qui a patiemment dactylographié le texte à caractère souvent technique.

TABLE DES MATIÈRES

page

REMERCIEMENTS..... i

TABLE DES MATIÈRES..... iii

INTRODUCTION..... 1

I EXPOSÉ DES MÉTHODES D'ANALYSE FACTORIELLE..... 5

I.1 Bases de l'analyse factorielle..... 5

I.2 Les méthodes d'extraction..... 16

I.2.1 Les modèles..... 16

I.2.2 Les méthodes..... 18

I.2.2.1 L'analyse en composantes principales..... 18

I.2.2.2 La méthode des facteurs principaux..... 19

I.2.2.3 La méthode du maximum de vraisemblance..... 20

I.2.2.4 La méthode alpha..... 21

I.3 Les méthodes de rotation..... 22

I.3.1 L'interprétation des facteurs..... 22

I.3.2 Méthodes orthogonales..... 31

I.3.3 Méthodes obliques..... 32

I.3.3.1 La structure factorielle..... 32

I.3.3.2 La méthode promax..... 36

I.3.3.3 La méthode orthoblique..... 36

I.3.3.4 La méthode biquartimin..... 37

I.3.3.5 La méthode maxplane..... 38

I.4 L'indétermination des solutions..... 38

II ÉTUDE COMPARATIVE DE DIVERSES MÉTHODES..... 43

II.1 La position du problème..... 43

II.2 La procédure employée..... 44

II.2.1 Les types de matrices de corrélation..... 44

II.2.2 Les méthodes d'extraction et de rotation de l'étude..... 53

II.2.3 La liste des critères..... 54

II.3 Critère 1 : distance par rapport à la solution de compromis..... 57

II.3.1 Description..... 57

II.3.2 Le type I..... 59

II.3.3 Le type II..... 62

II.3.4 Le type III..... 65

II.3.5 Le type IV..... 65

II.3.6 Le type V..... 65

II.3.7 Le type VI..... 74

II.3.8 Le type VII..... 74

II.3.9 Le type VIII..... 81

II.3.10 Sommaire des résultats..... 81

II.4 Critère 2 : le temps d'exécution informatique..... 85

II.4.1 Temps relatif aux méthodes d'extraction..... 85

II.4.2 Temps relatif aux méthodes de rotation..... 88

II.4.3 Sommaire des résultats..... 90

II.5 Critère 3 : l'invariance factorielle..... 92

II.5.1 Description..... 92

II.5.2 Le type I..... 96

II.5.3 Le type II..... 98

II.5.4 Le type III..... 100

II.5.5 Le type IV..... 100
II.5.6 Le type V..... 100
II.5.7 Le type VI..... 100
II.5.8 Le type VII..... 106
II.5.9 Le type VIII..... 106
II.5.10 Sommaire des résultats..... 106

II.6 Critère 4 : le jugement des spécialistes..... 109

II.6.1 Le type I..... 109
II.6.2 Le type III..... 110
II.6.3 Le type V..... 110
II.6.4 Le type VII..... 111
II.6.5 Sommaire des résultats..... 115

II.7 Discussion et synthèse des résultats..... 117

II.8 Recommandations..... 125

CONCLUSION..... 127

BIBLIOGRAPHIE..... 129

ANNEXE I..... 135

ANNEXE II..... 139

INTRODUCTION

C'est au début du siècle que l'on situe l'origine de l'analyse factorielle suite aux travaux de Karl Pearson vers 1901 concernant les axes principaux et suite surtout à ceux de Charles Spearman vers 1904 sur la distinction entre facteurs communs et facteurs uniques.

Depuis lors, on a vu poindre une foule de développements théoriques de grande importance dont ceux de Spearman [1904] "two-factor theory", Hotelling [1933] "principal compoment analysis", Holzinger [1937] "bi-factor theory", Thurstone [1931] "multiple factor theory" et "simple structure theory", Kaiser [1958] "varimax rotation" et Joreskog [1969] "confirmatory factor analysis".

Mais qu'ont-ils apporté ces travaux en quatre-vingts ans ? A-t-on réussi à résoudre les problèmes inhérents à l'analyse factorielle, à en faire une méthode d'analyse fiable et aisément utilisable par les chercheurs en éducation et dans les domaines connexes comme la psychologie et la sociologie ?

En réalité tous ces travaux ont contribué à meubler considérablement l'arsenal méthodologique des chercheurs. De telle sorte que sous le vocable "analyse factorielle" se cache non pas une mais bien plusieurs méthodes d'analyse. Ce qui peut être un atout pour la science mais qui risque de poser de sérieux problèmes de choix d'une "méthode appropriée" aux utilisateurs de l'analyse factorielle, c'est-à-dire les chercheurs qui en font un usage modéré mais réfléchi.

Car il existe bien sûr deux issues plutôt radicales à ce problème de choix. La première consiste à laisser prendre les décisions au consultant, méthodologiste, statisticien ou informaticien qui a souvent au mieux une idée vague du champ d'expertise du chercheur. Ce consultant est parfois obligé d'effectuer des choix très rapides sans connaître véritablement la provenance des données, la nature des variables en cause, ou même les objectifs de la recherche. Thorndike [1978, p. 322] souligne les dangers de cette façon de procéder qu'il a appelé péjorativement "saturday-night-special factor analysis".

Une deuxième issue consiste à se spécialiser en analyse factorielle. Cependant cette issue n'est pas réaliste pour la très grande majorité des chercheurs en éducation, compte tenu de l'investissement de temps considérable que cela exige. Car même le chercheur que l'on peut qualifier de méthodologiste, c'est-à-dire celui qui s'intéresse à la mesure, la psychométrie, les statistiques, ne peut se permettre une spécialisation trop marquée en analyse factorielle au risque de devoir abandonner ses études dans la plupart des autres secteurs de la méthodologie de la recherche.

Qu'arrive-t-il alors des très nombreux utilisateurs "modérés" de l'analyse factorielle, ceux qui ne voulant pas devenir spécialistes, ont tout de même le souci d'effectuer une analyse plus pertinente qu'une "étude de samedi soir" ?

Aucune réponse unilatérale n'est possible bien sûr mais les pages qui vont suivre sont écrites en vue d'aider ceux qui, comme nous, se sont posés la question. Nous sommes d'avis qu'alors nous rejoindrons la très grande majorité des utilisateurs. Car ceux-ci sont nombreux comme l'ont déjà montré Anderson et alii [1977]. Ces auteurs ont en effet recensé qu'en 1966 plus de 1000 articles avaient employé l'analyse

factorielle en psychologie, éducation et dans les domaines connexes des sciences humaines. Dans une étude plus récente, Ajar [1982] dénombre, en date de 1981, 880 publications utilisant cette méthode dans le seul secteur de la mesure et de l'évaluation. Harman [1976] souligne des applications de l'analyse factorielle dans des domaines aussi variés que l'économique, la météorologie, les sciences politiques, l'urbanisation, l'archéologie et la médecine.

Il est donc raisonnable d'avancer, chiffres à l'appui, qu'encore aujourd'hui l'analyse factorielle constitue une des méthodes d'analyse les plus utilisées. Harman [1976], Lawley et Maxwell [1971] et Tatsuoka [1971] avaient déjà fait cette remarque, il y a quelques années.

Ceci n'est pas si étonnant, compte tenu qu'il n'y a pas véritablement de procédure équivalente ou alternative à l'analyse factorielle. Le problème n'est donc pas de se demander si l'analyse factorielle doit être utilisée ou pas mais comment elle doit l'être. Plus précisément comment faire pour trouver la méthode d'analyse factorielle qui convienne à tel type de données ?

La deuxième partie de cet ouvrage abordera cette question d'un point de vue empirique en présentant une étude comparative de diverses méthodes d'analyse factorielle.

Mais avant d'entrer à fond dans cette problématique nous jugeons utile de faire une présentation sommaire, visuelle et intuitive, de l'analyse factorielle. C'est ce qui constitue la première partie du texte.

I EXPOSÉ DES MÉTHODES D'ANALYSE FACTORIELLE

I.1 Bases de l'analyse factorielle

L'analyse factorielle peut être décrite comme un ensemble de méthodes servant soit à classer les variables en tenant compte de leurs associations (ou corrélations), soit à classer les sujets en tenant compte des relations entre leurs scores à diverses variables (Reuchlin [1964, p.13]).

La plupart du temps le point de départ de l'analyse sera une matrice de corrélations entre les "n" variables observées. On en arrivera à construire d'autres variables (hypothétiques) appelées facteurs. Généralement, on ne retiendra que m facteurs où m n, le but étant d'obtenir une description plus parcimonieuse que celle de départ sans perte substantielle d'information.

À titre d'illustration, regardons les données du tableau 1. Elles représentent les résultats de 25 élèves à un test de mathématiques et à un test de physique. On remarque que même si les résultats sont plus élevés en mathématiques, les meilleurs élèves en mathématiques sont aussi les meilleurs en physique. Une association positive relativement forte existe donc entre les deux variables observées :

X = test en mathématiques
Y = test en physique

Le coefficient de corrélation s'élève en effet à 0,83.

Comme il arrive souvent que l'on utilise les scores centrés et réduits en analyse factorielle, nous les avons aussi reportés au tableau 1. Rappelons que

$$Z_X = \frac{X - m_X}{e_X}, \quad Z_Y = \frac{Y - m_Y}{e_Y}$$

où m_X, m_Y sont des moyennes et e_X, e_Y des écarts-types.

Le graphe 1 montre que les points se regroupent pour former un ovale ou ellipse. Les deux axes principaux de cette ellipse sont des facteurs, au sens de l'analyse en composantes principales, dont l'importance est proportionnelle à leur longueur. Plus l'axe est grand, plus il rend compte de la relation r_{XY} ou de la variance r^2_{XY} entre les variables, donc de la configuration des points. À l'aide du graphe 2, qui utilise la représentation centrée et réduite de cette ellipse, fréquemment employée par les spécialistes du domaine, nous allons expliquer le sens de cette dernière assertion.

Nous voyons qu'en général la position d'un point (x, y) est mieux estimée par le point correspondant (Fx, Fy) situé sur le premier axe principal F que par le point correspondant (G_x, G_y) situé sur le deuxième axe principal G. Ceci parce que G est plus court que F et donc les distances comme d_F entre (x, y) et (F_x, F_y) seront plus courtes que les distances comme d_G entre (x, y) et (G_x, G_y).

C'est-à-dire que l'axe F est plus près du nuage de points au sens des moindres carrés. Plus le lien linéaire entre X et Y augmente, plus G rétrécit et meilleure est la qualité de la représentation des points (x, y) par les points

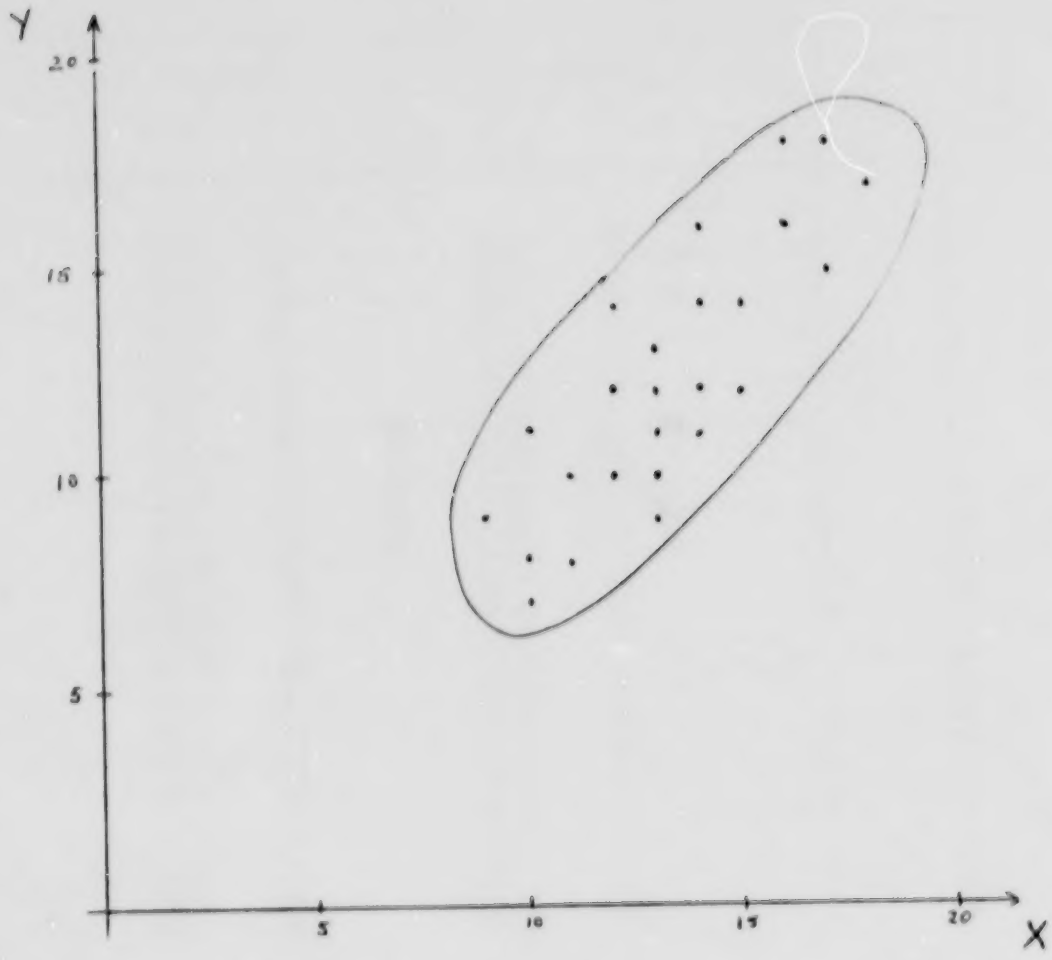
TABLEAU 1 : Scores de 25 élèves en mathématiques et en physique

Élève \ Test	Mathématiques X	Z_X	Physique Y	Z_Y
1	10	-1,36	7	-1,67
2	10	-1,36	8	-1,35
3	11	-0,95	8	-1,35
4	9	-1,78	9	-1,04
5	13	-0,12	9	-1,04
6	11	-0,95	10	-0,72
7	12	-0,53	10	-0,72
8	13	-0,12	10	-0,72
9	10	-1,36	11	-0,41
10	13	-0,12	11	-0,41
11	14	0,30	11	-0,41
12	12	-0,53	12	-0,09
13	13	-0,12	12	-0,09
14	14	0,30	12	-0,09
15	15	0,71	12	-0,09
16	13	-0,12	13	0,23
17	12	-0,53	14	0,54
18	14	0,30	14	0,54
19	15	0,71	14	0,54
20	17	1,54	15	0,86
21	14	0,30	16	1,18
22	16	1,13	16	1,18
23	18	1,96	17	1,49
24	16	1,13	18	1,81
25	17	1,54	18	1,81

$m_X = 13,28$ $\quad m_Y = 12,28$

$e_X = 2,41$ $\quad e_Y = 3,16$

$r_{XY} = 0,83$



Graphe 1 : Représentation géométrique des scores de 25 élèves en mathématiques et en physique

(F_x, F_y) situés sur le premier axe principal. À la limite, lorsque la corrélation entre X et Y est parfaite ($r_{XY} = 1$) l'ellipse se confond en une droite qui est justement l'axe principal F. La qualité de la représentation des points (x, y) est donc parfaite dans ce cas, les distances d_F étant nulles.

Dans l'exemple qui nous intéresse, on peut montrer que l'axe F explique 91,7 pour cent de la variance totale des variables alors que l'axe G n'explique que 8,3 pour cent (voir tableau 2). La procédure habituelle dans un pareil cas consiste à ne tenir compte que du premier axe principal, i.e. le facteur F. Le facteur G est mis de côté puisqu'il n'explique pas suffisamment de variance. On a alors atteint un objectif de parcimonie: les deux variables de départ, X et Y pouvant être expliquées suffisamment par un facteur sans perte véritable d'information. On a l'habitude de nommer un facteur ainsi obtenu en tenant compte des variables qu'il explique. Comme les variables X et Y ont trait respectivement à un test de mathématiques et à un test de physique, le nom de "sciences physiques" serait bien approprié au facteur F.

Grâce aux données du tableau 2, on obtient, compte tenu du modèle linéaire propre à l'analyse en composantes principales,

$$X = 0{,}95759F - 0{,}28812G \qquad (1)$$

$$Y = 0{,}95759F + 0{,}28812G \qquad (2)$$

et

$$F = 0{,}52214X + 0{,}52214Y \qquad (3)$$

$$G = -1{,}73539X + 1{,}73539Y \qquad (4)$$

Ainsi les scores pour les 25 étudiants au test hypothétique F de sciences physiques peuvent être calculés à l'aide de l'équation (3) et des scores des étudiants aux

TABLEAU 2 : Pattern factoriel et coefficients des scores factoriels selon l'analyse en composantes principales[1]

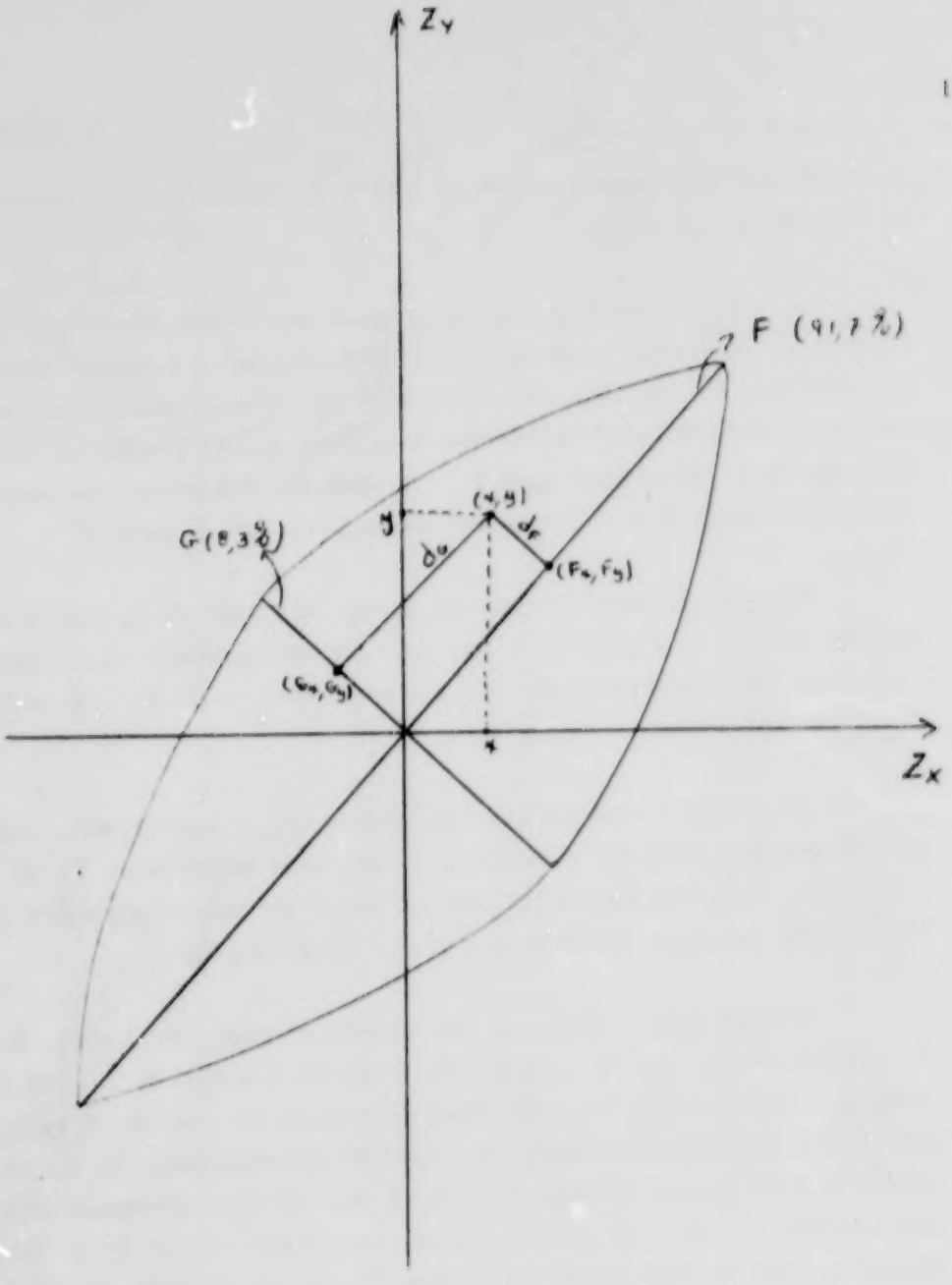
Pattern factoriel

	F	G
X	0,95759	-0,28812
Y	0,95759	+0,28812
Variance du facteur	1,83397	0,16603
Pourcentage de variance	91,7	8,3

Coefficients des scores factoriels

	F	G
X	0,52214	-1,73539
Y	0,52214	+1,73539

1 L'appendice I fournit les principales étapes de calcul de ces résultats.



Graphe 2 : Représentation géométrique des axes principaux F et G.

tests de mathématiques X et de physique Y. Les résultats apparaissent au tableau 3.

Il est à remarquer que les scores des élèves au facteur F constituent un compromis acceptable pour représenter les scores au test de mathématiques et au test de physique. Les élèves ayant bien réussi à ces deux derniers tests obtiennent aussi des scores élevés au test hypothétique de sciences physiques F. De même les élèves qui ont moins bien réussi aux tests X et Y obtiennent des scores bas au test F.

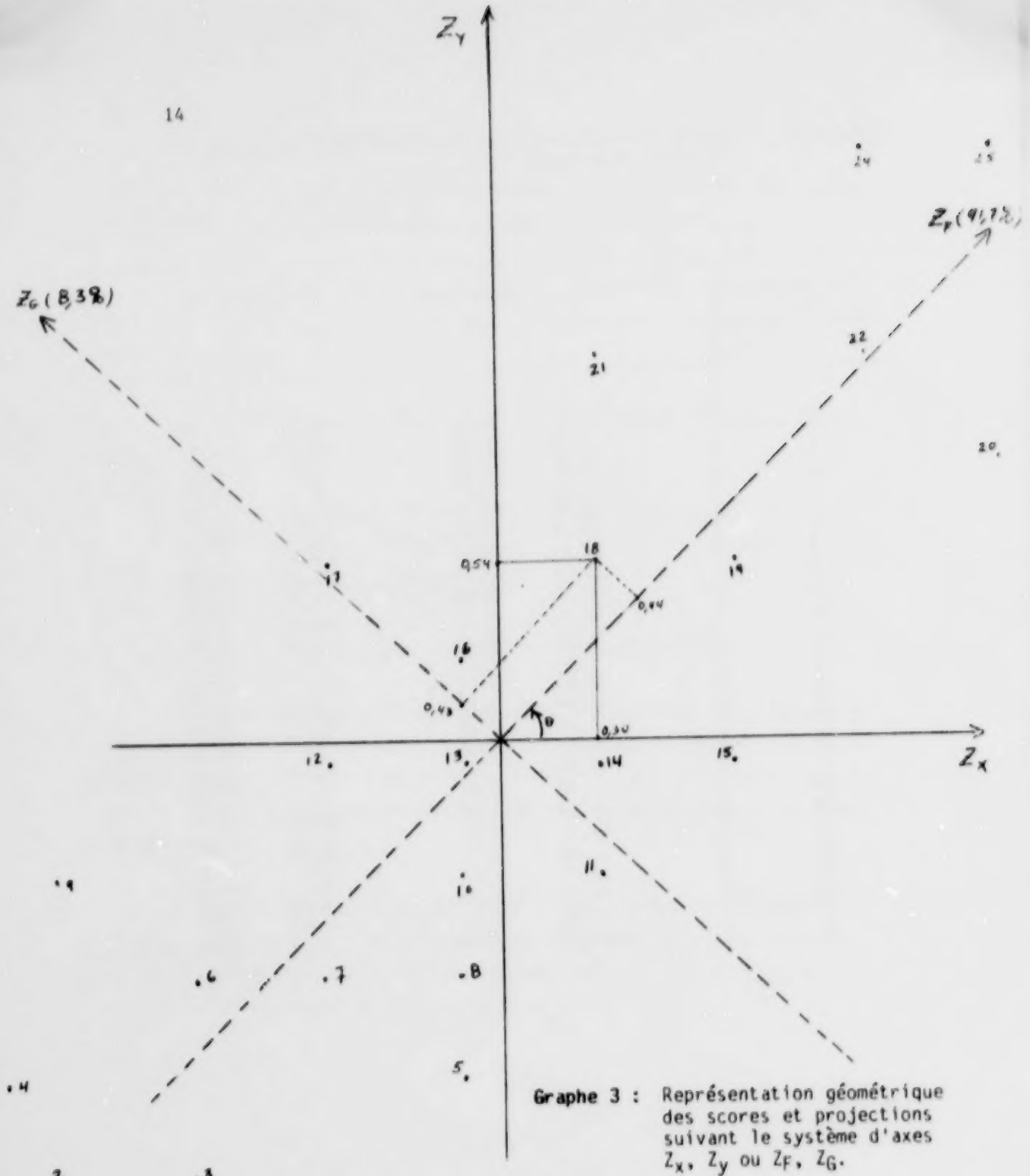
Nous avons donné les scores factoriels pour le facteur G afin de montrer qu'ils ne constituent pas du tout un compromis acceptable pour illustrer les résultats des élèves aux tests X et Y. Un bref coup d'oeil suffit pour s'en convaincre.

Le graphe 3 représente les points du graphe 1 mais une fois qu'ils ont été centrés et réduits. Les axes principaux Z_F et Z_G de l'ellipse ont été tracés en vue de vérifier les coordonnées de ces points dans les deux systèmes d'axes Z_X, Z_Y et Z_F, Z_G.

Tout se passe comme si, les points n'ayant pas bougé, les axes de référence Z_X, Z_Y ont subi une rotation d'angle θ. Prenons par exemple l'élève à qui on a attribué le numéro 18. Selon le tableau 1, ses scores centrés et réduits au test de mathématiques et au test de physique sont respectivement 0,30 et 0,54. Ce qui correspond bien aux projections respectives de ce point sur l'axe Z_X et sur l'axe Z_Y. Toujours pour le même élève ses scores factoriels centrés et réduits au facteur F de sciences physiques et au

TABLEAU 3 : Scores factoriels des 25 étudiants pour le facteur F "sciences physiques"

Élève \ Facteur	Sciences physiques F	Z_F	non interprété G	Z_G
1	8,88	-1,58	-5,21	-0,54
2	9,40	-1,41	-3,47	+0,11
3	9,92	-1,20	-5,21	-0,71
4	9,40	-1,47	0	+1,28
5	11,49	-0,61	-6,94	-1,60
6	10,96	-0,87	-1,74	+0,39
7	11,49	-0,65	-3,47	-0,33
8	12,01	-0,44	-5,21	-1,05
9	10,96	-0,92	+1,74	+1,66
10	12,53	-0,27	-3,47	-0,50
11	13,05	-0,06	-5,21	-1,22
12	12,53	-0,32	0	+0,77
13	13,05	-0,11	-1,74	+0,05
14	13,58	0,11	-3,47	-0,67
15	14,10	0,33	-5,21	-1,39
16	13,58	0,06	0	+0,60
17	13,58	0,01	+3,47	+1,87
18	14,62	0,44	0	+0,43
19	15,14	0,66	-1,74	-0,29
20	16,71	1,26	-3,47	-1,19
21	15,66	0,77	+3,47	+1,53
22	16,71	1,21	0	+0,08
23	18,27	1,80	-1,74	-0,81
24	17,75	1,54	+3,47	+1,18
25	18,27	1,75	+1,74	+0,46



Graphe 3 : Représentation géométrique des scores et projections suivant le système d'axes Z_x, Z_y ou Z_F, Z_G.

facteur G "non interprété" sont respectivement 0,44 et 0,43. Ces scores correspondent aux projections sur les axes respectifs Z_F et Z_G.

Il est possible enfin de reconstituer les scores aux tests X et Y à l'aide des scores factoriels des facteurs F et G au tableau 3 et à l'aide des équations ① et ②.

Par exemple pour l'élève auquel on a attribué le numéro 2, son score X reconstitué est bien 10 car

$$[0,95759 \times 9,4] + [(-0,28812) \times (-3,47)] = 10,$$

et son score Y reconstitué est bien 8 car

$$[0,95759 \times 9,4] + [(0,28812) \times (-3,47)] = 8.$$

Cet exemple, même s'il est simplifié à l'extrême, nous permet de constater comment on parvient à réduire les dimensions descriptives d'un domaine. À partir de n (ici n = 2) variables, on se retrouve avec m (ici m = 1) variable(s) ou facteur(s), où $m<n$, avec un minimum de perte d'information.

Or il n'existe pas qu'une seule façon de réduire le nombre de variables d'un domaine. Depuis Spearman, plusieurs méthodes d'extraction de facteurs ont été proposées et chacune avec ses particularités propres. C'est le sujet de la prochaine section.

I.2 Les méthodes d'extraction

I.2.1 Les modèles

Nous décrirons en premier lieu les deux grands modèles qui servent à l'extraction des facteurs.

Le modèle des composantes principales (A.C.P.) a comme objectif d'expliquer ou d'extraire le maximum de variance dans les données. Alors que le modèle classique de l'analyse factorielle (A.F.C.) vise à reproduire au mieux les corrélations de départ.

Comme l'exprime si justement Reuchlin [1964, p. 23], l'A.C.P. vise la meilleure reconstruction des données tandis que l'A.F.C. vise la meilleure reconstruction des corrélations.

Algébriquement, le modèle de l'A.C.P. se traduit par la série d'équations linéaires suivantes :

$$Z_1 = a_{11}F_1 + a_{12}F_2 + \dots + a_{1n}F_n$$

$$Z_2 = a_{21}F_1 + a_{22}F_2 + \dots + a_{2n}F_n$$

$$\vdots$$

$$Z_n = a_{n1}F_1 + a_{n2}F_2 + \dots + a_{nn}F_n.$$

où les Z_j sont les "n" variables observées,
les F_i sont les "n" facteurs (ou composantes principales), et les a_{ji} sont les saturations.

Un exemple d'une telle série d'équations linéaires est donné à la section 2 avec les équations ① et ② , où $n = 2$.

De même on peut représenter le modèle de l'A.F.C. par

$$Z_1 = a_{11}F_1 + a_{12}F_2 \text{ ------ } + a_{1m}F_m + b_1U_1$$
$$Z_2 = a_{21}F_1 + a_{22}F_2 + \text{ ------ } a_{2m}F_m + b_2U_2$$
$$\vdots$$
$$Z_n = a_{n1}F_1 + a_{n2} + F_2 + \text{ ------ } + a_{nm}F_m + b_nU_n$$

où les Z_j sont les "n" variables observées,
les F_i sont les "m" facteurs communs, où $m<n$,
les U_j sont les "n" facteurs dits uniques ou spécifiques,
les a_{ji} sont les saturations des facteurs communs,
les b_j sont les saturations des facteurs uniques.

Ainsi la différence de base entre les deux modèles, si on se réfère aux deux systèmes d'équations décrits plus haut, réside au niveau de la distinction que fait le modèle de l'A.F.C. entre facteurs communs et uniques. Cette distinction n'existe pas pour le modèle de l'A.C.P. Dans le cas de l'A.C.P., le modèle postule que chaque variable est décrite par "n" facteurs ; alors que pour l'A.F.C. chaque variable est décrite par "m + 1" facteurs.

L'objectif premier d'une analyse factorielle consiste à trouver les saturations a_{ji} à partir d'une matrice de corrélations entre les variables observées Z_j, et ce pour l'un comme pour l'autre modèle.

I.2.2 Les méthodes

Ces deux modèles donnent lieu à plusieurs méthodes d'extraction de facteurs, c'est-à-dire plusieurs méthodes menant aux saturations a_{ji} à partir d'une même matrice de corrélations.

I.2.2.1 L'analyse en composantes principales

Le modèle de l'A.C.P. ne permet qu'une méthode d'extraction, appelée l'analyse en composantes principales. Elle consiste à extraire "n" facteurs (ou composantes) à partir d'une matrice de corrélations entre les "n" variables Z_j. Géométriquement ces facteurs correspondent aux axes principaux d'un ellipsoïde à n dimensions, approximant la forme du nuage de points. Le graphe 3 de la section 2 donne un exemple de ce qu'on peut obtenir lorsqu'il y a $n = 2$ variables. Selon cette méthode, le premier facteur est celui qui contribue le plus à expliquer la variance des données ; le deuxième facteur est celui qui contribue le plus à expliquer la variance résiduelle des données, une fois l'effet du premier facteur enlevé ; le troisième facteur est celui qui contribue le plus à expliquer la variance résiduelle, une fois l'effet des deux premiers facteurs enlevé ; et ainsi de suite pour les autres facteurs.

En général, on ne retient que les "m" premiers facteurs d'une A.C.P., où $m < n$; à savoir ceux qui expliquent le maximum de variance. On atteint ainsi un but parcimonieux. La section 2 donne un exemple où $n = 2$, $m = 1$.

Du modèle classique sont issues plusieurs méthodes, toutes plus ou moins différentes les unes des autres. Nous décrivons ici trois méthodes qui ont chacune un trait bien distinctif.

1.2.2.2 La méthode des facteurs principaux

Cette méthode suit essentiellement la même procédure mathématique que l'A.C.P. mais avec la grande distinction qu'elle utilise une matrice de corrélations modifiée. Afin d'expliquer cette modification, il convient d'introduire un nouveau concept. La communauté h_j^2 d'une variable Zj est définie par :

$$h_j^2 = a_{j1}^2 + a_{j2}^2 + \cdots a_{jm}^2,$$

et exprime la contribution des m facteurs communs à la variance de Z_j. Comme les variables Z_j sont centrées (moyenne = 0) et réduites (variance = 1) on a (Harman [1976, p. 18]) :

$$h_j^2 + b_j^2 = 1 \text{ pour le modèle de l'A.F.C.}$$

La méthode des facteurs principaux utilise comme point de départ une matrice de corrélations dont les "1" de la diagonale principale doivent être remplacés par les communautés h_j. Car les facteurs que l'on veut obtenir sont ceux qui proviennent des relations communes entre les variables. L'influence des corrélations de "1", d'une variable avec elle-même, qui forment la diagonale d'une matrice de corrélations, doit être éliminée. (Rummel [1970], p. 314]). Or les communautés ne peuvent être calculées qu'une fois les facteurs communs trouvés. Il est donc essentiel d'obtenir une estimation "a priori" de ces communautés. Parmi les estimés les plus connus on peut mentionner,

MAXROW : la corrélation maximale pour chaque ligne de la matrice de corrélations, excluant la diagonale principale,
et SMC : le carré de la corrélation multiple.

Cette méthode est très générale et englobe en quelque sorte l'A.C.P.. En effet, pour le modèle de l'A.C.P. les b_j^2 sont nuls et $h_j^2 = 1$. Ainsi aucune modification de la diagonale de la matrice de corrélations n'est requise dans ce cas.

(Voir Harman [1976, p. 133], Kim et Mueller [1978b, p. 14]).

I.2.2.3 La méthode du maximum de vraisemblance

Comme pour toutes les méthodes du modèle classique, la méthode du maximum de vraisemblance a pour objectif de choisir les saturations a_{ji} qui reproduisent au mieux les corrélations observées. Ces saturations seront, dans ce cas-ci, des estimateurs (au sens du critère statistique du maximum de vraisemblance) des saturations, vues en tant que paramètres, d'une population de laquelle l'échantillon des données a été pris. Cette méthode exige une estimation préalable du nombre de facteurs. On doit assumer en outre que les variables suivent une distribution normale multidimensionnelle. Un test de signification (khi-deux) est utilisé ensuite pour vérifier l'adéquation du modèle postulé quant au nombre de facteurs.

En pratique, si le chercheur n'a aucune idée du nombre de facteurs, la procédure suivante peut être employée. On teste tout d'abord l'hypothèse d'une solution à un facteur commun. Si le test est significatif, on augmente à deux facteurs communs. Ainsi de suite jusqu'à ce que le test soit non significatif. Bien entendu il faut agir avec prudence lors de l'interprétation de ce test de signification sur le nombre de facteurs. La taille de l'échantillon peut en effet être un élément déterminant quant au rejet de telle ou telle hypothèse relative au nombre de facteurs. Si la taille de l'échantillon est très grande par exemple, un trop grand nombre de facteurs peut être retenu.

Lorsqu'il n'y a pas de test de signification disponible, une autre procédure doit être utilisée. Il s'agit d'employer un estimé initial de communautés au lieu d'un estimé du nombre de facteurs : le carré de la corrélation multiple est un choix adéquat dans ce cas. (Voir Harman [1976, p. 197], Reuchlin [1964, p. 142], Kim et Mueller [1978b, p. 23]).

I.2.2.4 La méthode alpha

Alors que la méthode du maximum de vraisemblance est pertinente à l'inférence statistique au niveau d'une population de sujets, la méthode alpha permet une inférence psychométrique au niveau des variables. C'est-à-dire que les variables utilisées dans l'analyse sont considérées comme provenant d'une population de variables.

Les saturations seront choisies de telle sorte que les facteurs communs de l'échantillon de variables obtiennent une corrélation maximale avec les facteurs communs (hypothétiques) de la population des variables.

La procédure mathématique consiste à maximiser le coefficient alpha de Cronbach pour chaque facteur. Un estimé initial des communautés doit être fourni : le carré de la corrélation multiple est souvent employé à cette fin. Les facteurs retenus seront ceux pour lesquels le coefficient alpha est positif. Ce qui revient à ne retenir que les facteurs dont la valeur propre est supérieure à 1. (Voir Harman [1976, p. 228], Kim et Mueller [1978b, p. 26]).

I.3 Les méthodes de rotation

I.3.1 L'interprétation des facteurs

Une fois les facteurs extraits, il convient de les interpréter au mieux. Par exemple à la section 2, nous avons nommé le facteur F "sciences physiques" parce que les saturations du facteur F étaient élevées pour les deux variables X (mathématiques) et Y (physique). Or le pattern factoriel est rarement aussi simple. Deux exemples typiques de pattern factoriel se trouvent au tableau 4. Dans les deux cas, les saturations ont été obtenues à l'aide de la méthode de l'analyse en composantes principales (A.C.P.). On voit que :

1) les saturations du facteur 1 sont beaucoup plus importantes que celles du facteur 2 (ce qui d'ailleurs est une caractéristique d'une A.C.P.)

2) plusieurs variables donnent lieu à des saturations élevées pour les deux facteurs.

3) le facteur 2 est bipolaire, c'est-à-dire qu'il a des saturations élevées positives et négatives.

TABLEAU 4 a) : Saturations des deux premiers facteurs obtenues à l'aide d'une A.C.P. pour l'exemple de Reuchlin [1964, p. 10]

	Facteur 1	Facteur 2
Mathématiques	0,82	-0,57
Sciences	0,91	-0,42
Français	0,76	0,65
Latin	0,93	0,38

TABLEAU 4 b) : Saturations des deux premiers facteurs obtenues à l'aide à l'aide d'une A.C.P. pour l'exemple- type 1 des 12 variables du "Thurstone box problem".

	Facteur 1	Facteur 2
x	0,58	0,64
y	0,65	-0,57
x^2z	0,82	0,52
y^2z	0,85	-0,49
xz	0,92	0,38
$\sqrt{y^2+z^2}$	0,90	-0,38
$2x+2z$	0,91	-0,39
yz	0,91	0,38
xz^2	0,93	0,28
yz^2	0,93	-0,28
x/z	-0,21	0,49
y/z	-0,18	-0,56

Ces caractéristiques des facteurs, obtenus à l'aide d'une méthode d'extraction comme celles décrites à la section 3, nuisent la plupart du temps à une interprétation simple et adéquate des facteurs.

Or il existe un moyen de contrevenir à ce problème.

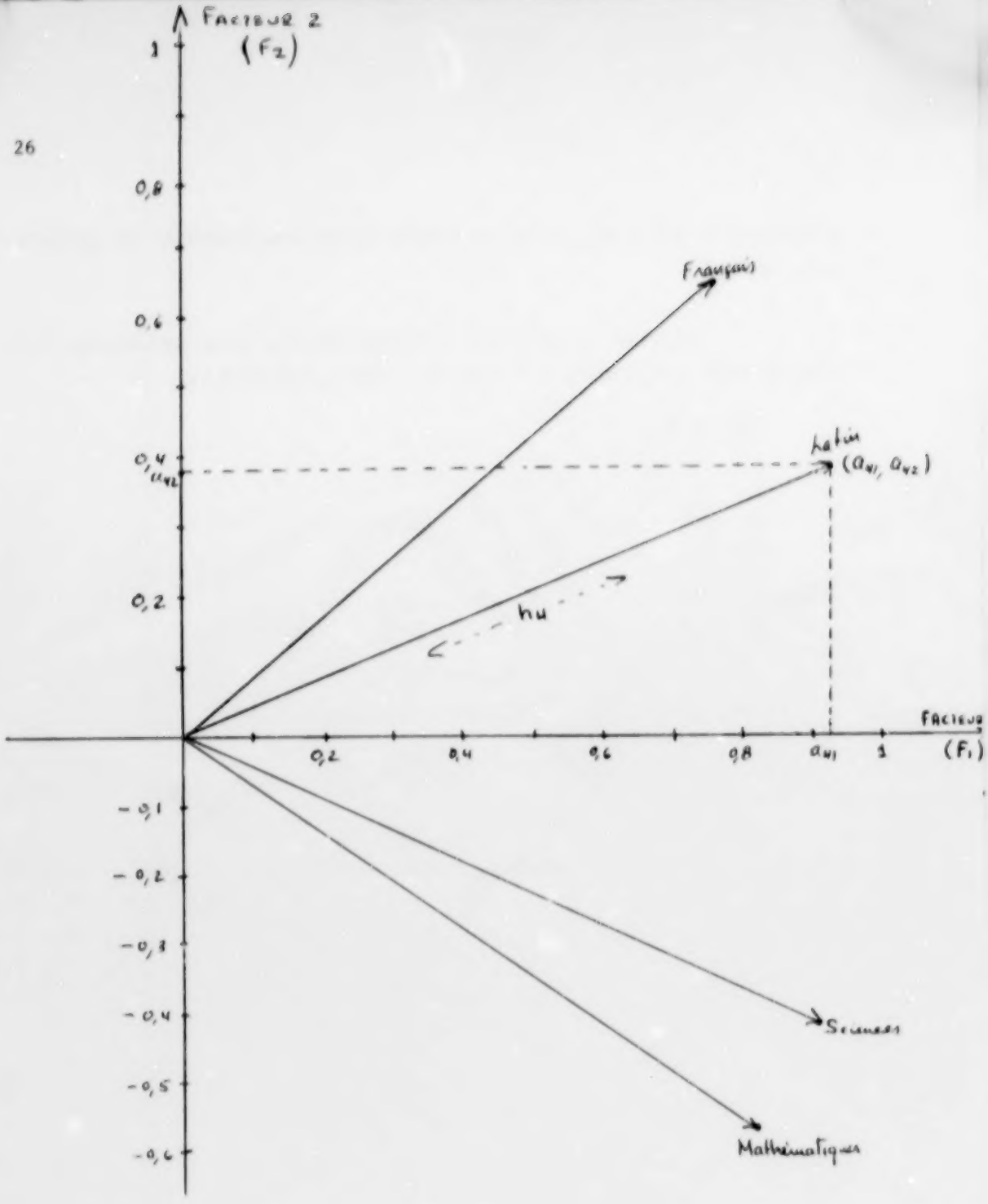
Mais avant d'aborder directement cette question, il importe de trouver une autre représentation géométrique des variables. Reportons-nous à la section 2. Le graphe 1 constitue une représentation des données du tableau 1 : les variables sont identifiées aux axes X (mathématiques) et Y (physique) et les individus aux points. Cette façon est peut-être la plus naturelle de représenter des données.

Mais comme en analyse factorielle on est souvent plus préoccupé par la "position" des variables (que des individus) les unes par rapport aux autres, la représentation géométrique qui suit nous permettra une vision plus claire du problème d'interprétation. Il s'agit tout simplement de considérer les variables comme des points et d'utiliser les facteurs comme axes, la position des variables les unes par rapport aux autres étant donnée par les saturations.

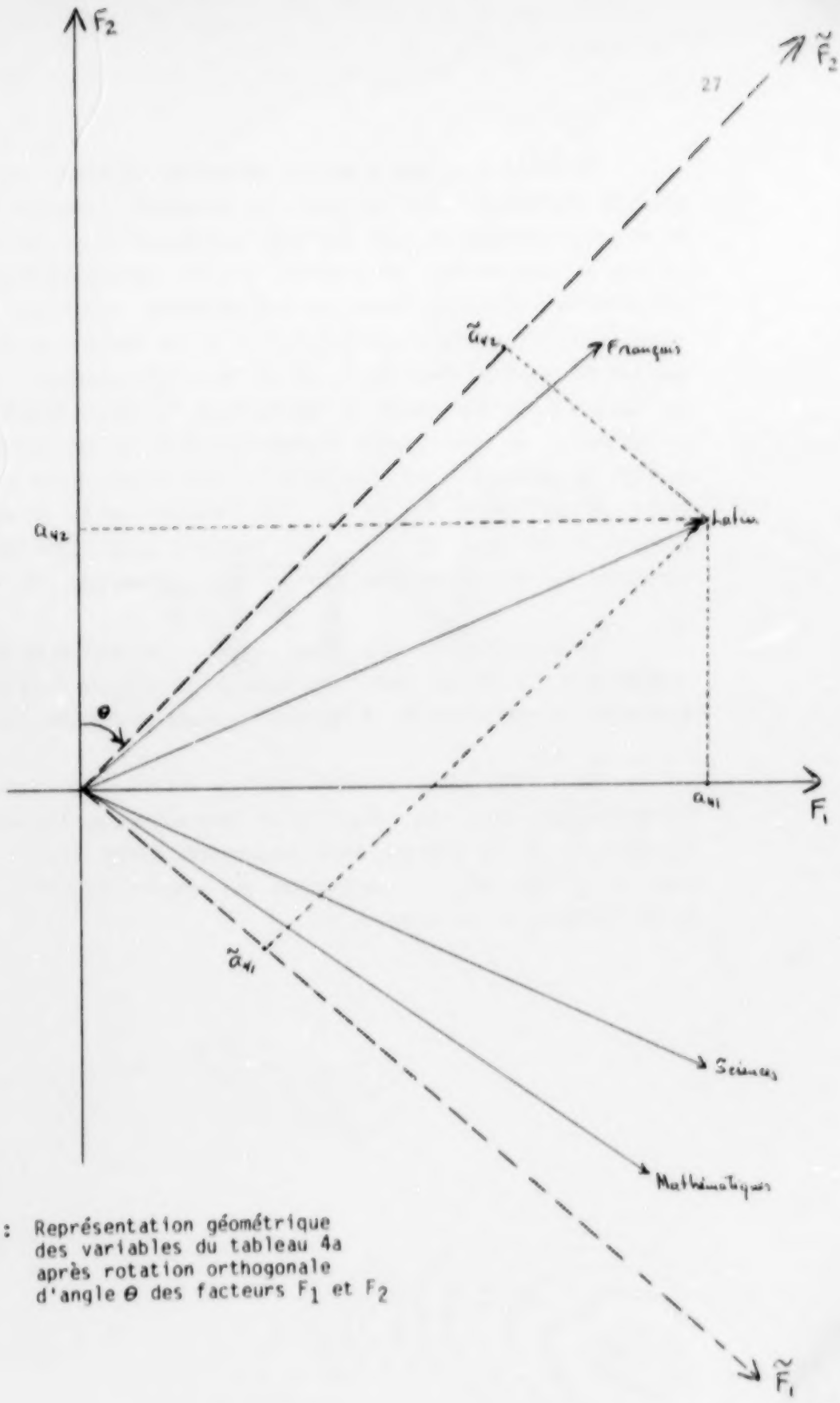
Le graphe 4 donne un exemple de ce que l'on peut obtenir en représentant de cette façon les variables et les facteurs du tableau 4a. Comme la longueur du vecteur qui part de l'origine et qui va jusqu'au point (variable) donne un indice de l'importance de cette variable pour la définition des facteurs, la coutume est de tracer ce vecteur pour chaque variable. Par exemple pour la variable "Latin" de coordonnées ($a_{41} = 0,93$, $a_{42} = 0,38$) la longueur du vecteur est $\sqrt{a_{41}^2 + a_{42}^2} = 1$ (théorème de Pythagore). Or cette longueur, tel qu'on l'a indiqué sur le

graphe n'est autre que la racine carrée de la communauté de la variable, soit h_4.

La solution au problème d'interprétation peut maintenant être exposée plus visiblement à l'aide de cette représentation.



Graphe 4 : Représentation géométrique des variables du tableau 4a en utilisant les facteurs comme axes.



Graphe 5 : Représentation géométrique des variables du tableau 4a après rotation orthogonale d'angle θ des facteurs F_1 et F_2

On voit, au graphe 4 que les variables "Sciences" et "Mathématiques" se regroupent ; de même pour les variables "Français" et "Latin". Or ces regroupements ne sont pas très manifestes si on se fie aux saturations du tableau 4a. On aimerait obtenir une représentation de ces variables qui réflètent mieux ces regroupements. Pour ce, on fera une rotation des axes (facteur 1 et facteur 2) de manière à ce que chaque axe (ou facteur) se rapproche d'un des deux regroupements : par exemple le facteur 1 se rapprochant du regroupement "Sciences-Mathématiques" et le facteur 2 du regroupement "Français-Latin", un peu comme au graphe 5. Ici la rotation θ est orthogonale, c'est-à-dire qu'on a toujours un angle de 90° entre $\tilde{F}_1$ et $\tilde{F}_2$, les facteurs après rotation, comme c'était le cas pour F_1 et F_2, les facteurs avant rotation. Les saturations $\tilde{a}_{ji}$ de ces nouveaux facteurs sont reproduites au tableau 5.

La situation est alors plus claire. On peut plus aisément interpréter à l'aide des nouvelles saturations $\tilde{a}_{ji}$ le facteur $\tilde{F}_1$ comme un facteur scientifique et le facteur $\tilde{F}_2$ comme un facteur littéraire.

On a fait subir le même type de rotation aux facteurs F_1 et F_2 du tableau 4b. Les résultats se retrouvent au tableau 5b. Les facteurs $\tilde{F}_1$ et $\tilde{F}_2$ peuvent être interprétés comme suit. Les variables x, y, x^2z, y^2, ... constituent des mesures de boîtes de hauteur x, de largeur y et de longueur z.

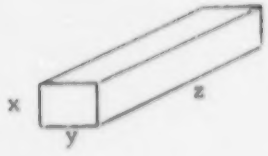


TABLEAU 5 a) : Saturations des facteurs du tableau 4a après rotation varimax.

	$\vec{F}_1$	$\vec{F}_2$
Mathématiques	0,99	0,15
Sciences	0,95	0,32
Français	0,11	0,99
Latin	0,42	0,91

TABLEAU 5 b) : Saturations des facteurs du tableau 4b après rotation varimax.

	$\vec{F}_1$	$\vec{F}_2$
x	0,01	0,86
y	0,87	0,01
x^2z	0,27	0,94
y^2z	0,96	0,20
xz	0,43	0,90
$\sqrt{y^2 + z^2}$	0,93	0,32
yz	0,94	0,31
$2x + 2z$	0,43	0,89
xz^2	0,50	0,83
yz^2	0,88	0,41
x/z	-0,49	0,22
y/z	0,23	-0,54

Puisque les saturations élevées sur le facteur $\bar{F}_1$ correspondent surtout aux mesures comportant la dimension y et que les saturations élevées sur le facteur $\bar{F}_2$ correspondent surtout aux mesures comportant la dimension x, on pourra appeler $\tilde{F}_1$ le facteur "largeur" et $\tilde{F}_2$ le facteur "hauteur".

La solution au problème d'interprétation semble donc être d'effectuer des rotations pour les facteurs, les variables restant fixes.

Cependant cela ne règle pas la question définitivement. Car il reste à déterminer quelle(s) rotation(s) sera(ont) utilisé(es). Sur quels critères nous baserons-nous pour effectuer ces rotations ?

Le critère le plus reconnu de nos jours est celui de "structure simple" dû à Thurstone [1947]. Celui-ci a défini le critère de structure simple en décrivant cinq sous-critères plus ou moins opérationnels dont le plus connu (et utile) est : chaque ligne du pattern factoriel doit avoir au moins un zéro (pour des facteurs non correlés). Ce qu'il faut retenir en fait de ce critère, c'est que le but ultime consiste à obtenir des facteurs tel qu'à chaque variable corresponde une seule saturation importante. Bien sûr cet objectif idéal est rarement atteint. Pourtant on remarque que la solution du tableeau 5b, par exemple, tend vers cet idéal : la plupart des variables obtiennent une saturation élevée pour un seul facteur, la saturation pour l'autre facteur étant beaucoup plus faible (idéalement tendant vers zéro).

Néanmoins plusieurs méthodes de rotation, soit graphiques ou subjectives, soit analytiques ou objectives, sont basées sur ce critère. Nous ne décrivons ici que quelques-unes des méthodes analytiques.

I.3.2 Méthodes orthogonales.

Une rotation orthogonale de deux facteurs F_1 et F_2 consiste à faire pivoter ces facteurs (axes) d'un angle θ de sorte qu'un angle de 90° demeure entre les facteurs après la rotation, $\tilde{F}_1$ et $\tilde{F}_2$ (comme au graphe 5).

Une rotation orthogonale effectuée sur plus de deux facteurs (v.g. "m") peut être visualisée comme le résultat de plusieurs rotations orthogonales de toute combinaison de deux facteurs parmi les "m" retenus. (voir Harman [1976, p. 250]).

Les plus connues des méthodes de rotation orthogonale peuvent se regrouper sous le vocable de rotation "orthomax". La procédure consiste à maximiser une certaine fonction des saturations des "facteurs après rotation". Ainsi les saturations a_{ji} recherchées seront celles qui maximisent

$$\Omega = \sum_{i=1}^{m} \left[\sum_{j=1}^{n} \tilde{a}_{ji}^4 - \frac{\gamma}{n} \left(\sum_{j=1}^{n} a_{ji}^2 \right)^2 \right]$$

où $\tilde{a}_{ji}$: saturations des "facteurs après rotations"

n : nombre de variables

m : nombre de facteurs

γ : à déterminer.

Pour certaines valeurs de γ, on retrouve des méthodes bien connues comme varimax ($\gamma = 1$), quartimax ($\gamma = 0$), biquartimax ($\gamma = 0,5$), equamax ($\gamma = m/2$). (Voir Harman [1976, p. 299], Bertrand et Dufour [1981]).

I.3.3 Méthodes obliques

I.3.3.1 La structure factorielle

Une rotation oblique de deux facteurs F_1 et F_2 consiste à faire pivoter les facteurs (axes) sans restriction d'orthogonalité entre les facteurs $\widetilde{F}_1$ et $\widetilde{F}_2$ obtenus après la rotation.

Or afin d'étudier sans heurt ce type de rotation, il convient d'introduire un nouveau concept : la structure factorielle. On se souviendra à la section I.1 que le pattern factoriel (tableau 2) donne les équations de régression des variables en fonction des facteurs (cf. équations ① et ②). Dans le cas où ces facteurs sont orthogonaux, les coefficients (de régression) du pattern sont aussi des coefficients de corrélation entre les variables et les facteurs respectifs. Par exemple le coefficient 0,95759 est un coefficient de régression de l'équation ①, X = 0,95759F - 0,28812G, mais aussi un coefficient de corrélation entre la variable X et le facteur F[1].

---

1 Il est facile de vérifier ce résultat : il s'agit de mettre en corrélation la colonne X du tableau 1 et la colonne F du tableau 3.

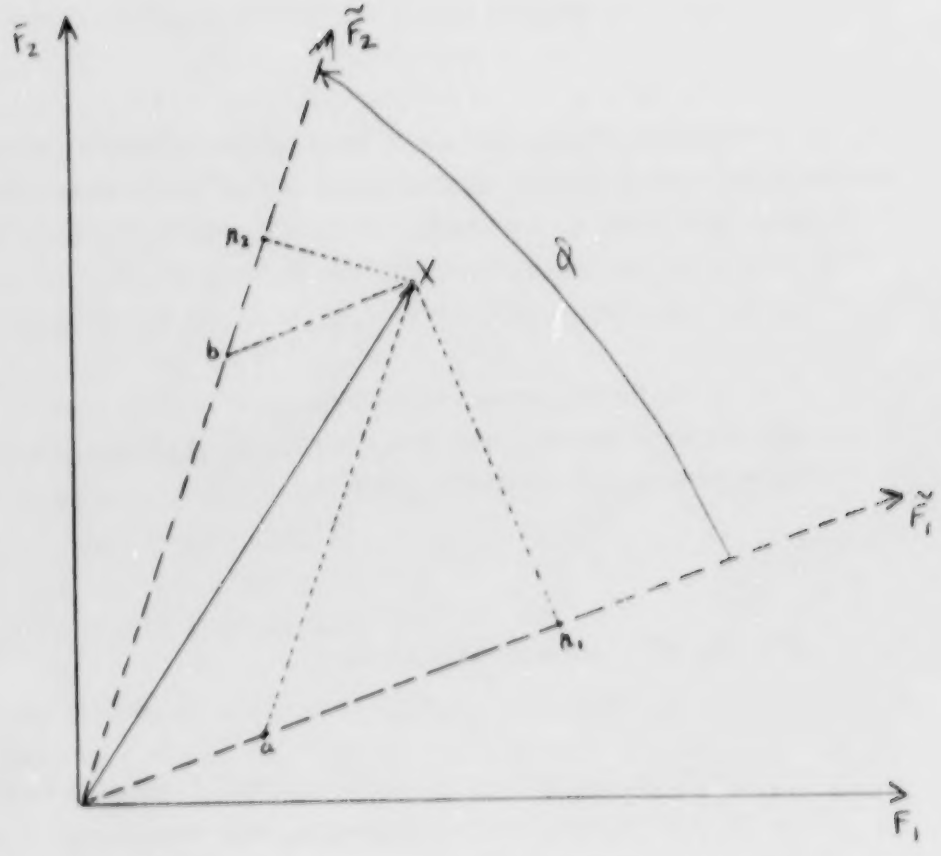
Mais si les facteurs ne sont pas orthogonaux, ce résultat n'est en général plus exact (Kim et Mueller [1978a, p. 26]). Il faut donc distinguer coefficients de régression et coefficients de corrélation lorsqu'on est aux prises avec des facteurs obliques. Ainsi le tableau des coefficients de régression est appelé pattern factoriel (comme au tableau 2) et celui des coefficients de corrélation, structure factorielle. Dans le cas de facteurs obliques ces deux tableaux sont essentiels pour décrire complètement la solution factorielle : la structure sert à identifier les variables associées à tel ou tel facteur et au calcul des scores factoriels, alors que le pattern est utilisé pour reproduire les corrélations initiales entre les variables et pour comparer différentes solutions (Harman [1976, p. 22]).

La représentation factorielle du graphe 6 nous aidera à visualiser la distinction pattern-structure (Rummel [1971, p. 398]).

On voit que les facteurs F_1 et F_2 ont subi une rotation oblique car l'angle ϕ , entre les facteurs après rotation, $\widetilde{F}_1$ et $\widetilde{F}_2$, n'est pas 90°. Considérons la variable X quelconque. On peut la projeter sur les axes-facteurs $\widetilde{F}_1$ et $\widetilde{F}_2$ de deux façons. Soit en faisant une projection oblique i.e. parallèle à $\widetilde{F}_1$ ou à F_2, dans lequel cas on obtient les coordonnées (a, b) pour la variable X ; soit en faisant une projection orthogonale sur $\widetilde{F}_1$ ou $\widetilde{F}_2$, i.e. d'angle de 90° avec $\widetilde{F}_1$ ou $\widetilde{F}_2$, ce qui donne (r_1, r_2) comme coordonnées. Remarquons alors que a et b sont les saturations du pattern factoriel, i.e. les coefficients de régression dans l'équation $X = a\,\widetilde{F}_1 + b\,\widetilde{F}_2$.

Cependant r_1 et r_2 sont les coefficients de corrélation de la structure factorielle.

On voit bien que si les facteurs $\tilde{F}_1$ et $\tilde{F}_2$ avaient été orthogonaux (en imaginant, par exemple, qu'ils s'identifient respectivement à F_1 et F_2), les projections obliques auraient été en fait des projections orthogonales. Dans ce cas, on aurait $a = r_1$, $b = r_2$ et donc le pattern s'identifierait à la structure.



Graphe 6 : Représentation géométrique d'une rotation oblique des facteurs F_1 et F_2 ; après la rotation l'angle α entre les facteurs $\tilde{F}_1$ et $\tilde{F}_2$ n'est pas 90°.

Dans ce qui suit, on retrouvera la description des quatre méthodes de rotation oblique parmi les plus utilisées.

I.3.3.2 La méthode promax

Cette méthode est due à Hendrickson et White [1964]. La procédure consiste à obtenir tout d'abord une solution orthogonale (habituellement quartimax ou varimax). Puis une solution idéale est construite en diminuant les saturations faibles et moyennes tout en maintenant élevées les saturations déjà élevées de la solution orthogonale.

La solution promax recherchée sera la plus ajustée à cette solution idéale au sens des moindres carrés, en utilisant une transformation dite de Procrustes[1] (Gorsuch [1974, p. 166]).

I.3.3.3 La méthode orthoblique

Initiée par Harris et Kaiser [1964], cette méthode utilise une série de transformations orthogonales pour en arriver à une solution oblique. Habituellement les critères quartimax ou varimax

---

1 Ce nom vient de la mythologie grecque. Un voleur de grand chemin, nommé Procrustes, forçait supposément toutes ses victimes à s'ajuster à son lit, étirant les plus petites et coupant les plus grandes ! (Voir Hurley et Cattell [1962]).

sont utilisés à cette fin (Gorsuch [1974, p. 204] et Harman [1976, p. 327]).

I.3.3.4 La méthode biquartimin

Elle date de 1957 et a été inspirée par Carroll [1957] qui l'a généralisée (Carroll [1960]) à une classe de méthodes nommée oblimin dont l'objectif est de minimiser la somme

$$B = \sum_{p<q=1}^{m} \left(n \sum_{j=1}^{n} \tilde{a}_{jp}^2 \, \tilde{a}_{jq}^2 - \gamma \sum_{j=1}^{n} \tilde{a}_{jp}^2 \sum_{j=1}^{n} \tilde{a}_{jq}^2 \right)$$

où n : nombre de variables

m : nombre de facteurs

$\tilde{a}_{jp}$: coefficient, dans la structure factorielle, de la variable j_p et du facteur p.

γ : variable à déterminer.

Pour $\gamma = 0$, on obtient la méthode quartimin ; pour $\gamma = 1$, la méthode covarimin ; et pour $\gamma = 0{,}5$, la méthode biquartimin. Ce qui fait de cette dernière méthode, un compromis entre covarimin qui tend à donner des solutions trop orthogonales et quartimin donnant des solutions trop obliques. (Harman [1976, p. 310]).

I.3.3.5 La méthode maxplane

L'objectif ici est de maximiser le nombre de saturations nulles ou presque nulles (Cattell et Muerle [1960], Eber [1966]). La méthode est itérative et procède en utilisant successivement chaque paire de facteurs (Gorsuch [1974, p. 198], Rummel [1971, p. 416]). On doit souligner que cette méthode est fort distincte des méthodes obliques ou orthogonales décrites plus haut. Alors que les autres méthodes sont dites analytiques, la méthode maxplane est classée comme topologique (Cattell et Khanna [1977, p. 178]).

I.4 L'indétermination des solutions

Si on définit une solution factorielle comme la donnée du pattern factoriel et de la structure factorielle obtenue à partir d'une matrice de corrélations, il est facile de se rendre compte que l'on peut dénombrer une quantité imposante de telles solutions en combinant les diverses méthodes d'extraction et de rotation.

Le problème est donc de choisir une combinaison adéquate : méthode d'extraction - méthode de rotation. Ce choix est en outre crucial car les solutions factorielles provenant des diverses combinaisons des méthodes peuvent être drastiquement différentes pour une même matrice de corrélations. Bien que, encore aujourd'hui, plusieurs utilisateurs semblent ne considérer que la très populaire combinaison

A.C.P. - varimax (Howarth [1977], Allen [1978], Coché et Meehan [1979]).

Or même en ne faisant varier que les méthodes d'extraction (ou de rotation) les résultats peuvent parfois s'avérer fort distincts.

À preuve, l'exemple fictif suivant. En corrélant neuf tests deux à deux, on obtient la matrice de corrélations:

	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9
V_1	1								
V_2	0,72	1							
V_3	0,63	0,56	1						
V_4	0,54	0,48	0,42	1					
V_5	0,45	0,40	0,35	0,30	1				
V_6	0,36	0,32	0,28	0,24	0,20	1			
V_7	0,27	0,24	0,21	0,18	0,15	0,12	1		
V_8	0,18	0,16	0,14	0,12	0,10	0,08	0,06	1	
V_9	0,09	0,08	0,07	0,06	0,05	0,04	0,03	0,02	1

Si on utilise l'analyse en composante principale pour extraire les facteurs de cette matrice, on obtient neuf facteurs, dont nous retenons les deux premiers, plus significatifs. Le premier facteur rend compte de 37 pourcent de la variance totale entre les variables et le deuxième facteur, onze pourcent.

Voici le pattern factoriel pour ces deux facteurs :

	Facteur 1	Facteur 2
V_1	0,88	-0,01
V_2	0,83	-0,02
V_3	0,76	-0,02
V_4	0,68	-0,02
V_5	0,59	-0,03
V_6	0,49	-0,03
V_7	0,38	-0,05
V_8	0,26	-0,09
V_9	0,13	0,99
Variance du facteur	37%	11%
Valeur propre	3,33	0,99

D'autre part, la méthode des facteurs principaux avec les estimés de communautés égaux à 0,81 pour V_1, 0,64 pour V_2, 0,49 pour V_3, 0,36 pour V_4, 0,25 pour V_5, 0,16 pour V_6, 0,09 pour V_7, 0,04 pour V_8, et 0,01 pour V_9 donne un résultat bien distinct. En effet, dans ce cas on obtient un seul facteur qui rend compte de 100 pourcent de la variance totale entre les variables et le pattern factoriel est :

Facteur 1

V_1	0,90
V_2	0,80
V_3	0,70
V_4	0,60
V_5	0,50
V_6	0,40
V_7	0,30
V_8	0,20
V_9	0,10

Variance du facteur 100%
Valeur propre 2,85

Cet exemple montre à quel point on doit être prudent dans le choix d'une méthode d'extraction de facteurs. Ajoutons que la même prudence est de mise pour le choix d'une méthode de rotation.

Ce problème de choix d'une combinaison adéquate "méthode d'extraction-méthode de rotation" est accentué par le fait qu'il n'existe pas de critère indéfectible qui permette de la choisir d'une façon relativement objective. En somme une partie importante d'arbitraire est laissée à l'utilisateur quant aux choix à effectuer. Il n'est donc pas étonnant de lire des commentaires sur la subjectivité de l'analyse factorielle (Kim et Mueller [1978a, p. 70]). Comme l'indiquent si bien Manners et Brush [1979], l'analyse factorielle requiert une bonne part de "jugement artistique".

Dans un tout récent volume dédié à l'analyse factorielle, Cureton et D'agostino [1983] n'hésitent pas à soutenir qu'encore aujourd'hui "l'analyse factorielle est partiellement une science mathématique et partiellement un art[1]".

Il n'existe donc pas de relation claire qui détermine, pour tel type de données, une méthode d'extraction ou de rotation à utiliser.

D'où le problème d'indétermination des solutions.

Même si des tentatives théoriques ont été proposées en vue de résoudre ce problème (Bertrand [1977]), il semble que nous ayons encore besoin d'une foule de recherches empiriques en vue de le circonscrire d'une façon satisfaisante.

La deuxième partie de cet ouvrage constitue un pas dans cette direction.

II ÉTUDE COMPARATIVE DE DIVERSES SOLUTIONS

II.1 La position du problème

La recherche décrite au cours des prochaines sections constitue une contribution empirique au problème décrit en I.4 d'indétermination des solutions causée par les choix relatifs aux méthodes d'extraction et de rotation.

Rappelons que le terme "solution" signifie ici le résultat (pattern factoriel et/ou structure factorielle) de l'application d'une méthode d'extraction suivie d'une méthode de rotation à une matrice de corrélations.

Il est à noter que nous n'avons pas abordé dans cette recherche les problèmes toujours bien actuels touchant au choix de l'estimé des communautés, du nombre de facteurs ou encore de l'estimé des scores factoriels.

Le lecteur intéressé au problème du choix des communautés ou du nombre de facteurs est référé à Bentler et Woodword [1980] et Ajar [1982]. Quant au problème d'indétermination des scores factoriels, les contributions suivantes sont suggérées au lecteur : Steiger [1979 a et b], Mc Donald et Mulaik [1979], Morris [1979], Laux et Tucker [1979], Steiger [1980].

II.2 La procédure employée

II.2.1 Les types de matrices de corrélations

Une foule d'études empiriques ont déjà traité de la comparaison entre différentes méthodes d'extraction de facteurs. Par exemple celles de Browne [1968], Humphreys et Ilgen [1969], Tucker, Koopman et Linn [1969], Harris et Harris [1971], Humphreys et Montanelli [1975], Conger et Stallard [1976], Lee et Comrey [1978], Jackson, Borgatta et Goldsmith [1979]. D'autres ont porté sur la comparaison entre diverses méthodes de rotation : Kaiser [1958], Eyman, Dingman et Meyers [1962], Butler [1964], Harris et Kaiser [1964], Gorsuch [1970], Dielman, Cattell et Wagner [1972], Hofmann [1978], Bertrand et Dufour [1981].

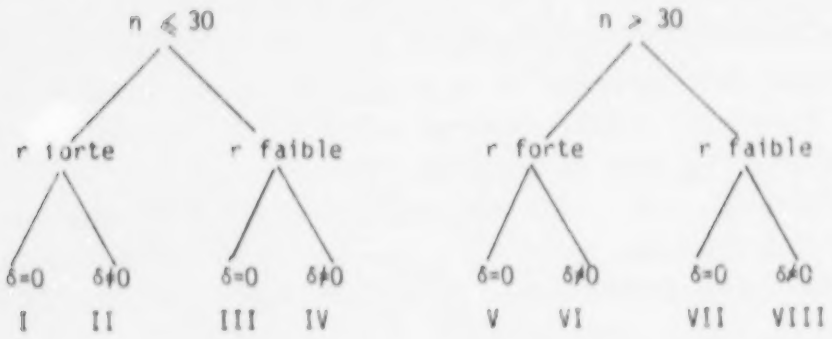
À partir de certains résultats de ces quelques recherches empiriques, nous avons établi huit types de matrices de corrélations. Ces types ont été définis à partir des trois considérations suivantes :

1) Le nombre de variables (n) de la matrice de corrélations : plus petit ou plus grand que 30.

2) L'intensité des corrélations (r) : elle peut être faible ou forte pour la majorité des corrélations de la matrice.

3) Le nombre de variables par groupe ou facteur : il sera identique ($\delta = 0$) ou différent ($\delta \neq 0$).

Le diagramme arborescent qui suit schématise les divers types.



Par exemple une matrice de corrélations du type I aura au plus 30 variables corrélées fortement et comprendra un nombre égal de variables par facteur. D'autre part une matrice de corrélations du type VIII aura au moins 30 variables corrélées faiblement et comprendra un nombre distinct de variables par facteur.

Pour chacun de ces types nous avons construit une matrice de corrélations respectant le plus possible les particularités du type choisi. Dans les paragraphes qui suivent nous décrivons ces matrices.

Dans le cas du type I, la matrice de corrélations est issue du "Thurstone box problem" (Thurstone [1947, p. 370]). Les variables retenues pour corrélation sont :

$V_1 : x$, $V_2 : y$

$V_5 : xz$, $V_6 : yz$

$V_9 : x^2z$, $V_{11} : y^2z$

$V_{10} : xz^2$, $V_{12} : yz^2$

$V_{15} : x/z$, $V_{17} : y/z$

$V_{20} : 2x + 2z$, $V_{24} : \sqrt{y^2 + z^2}$

Rappelons que chacune des variables constitue une mesure d'une boîte de dimensions x, y et z, tel qu'illustré à la section I.3. Ces variables ont été choisies parce qu'elles sont pour la plupart corrélées avec une intensité relativement forte et qu'elles se regroupent en deux dimensions (facteurs), chacune ayant un nombre égal de variables : V_1, V_5, V_9, V_{10}, V_{15}, V_{20} pour la dimension "x" et V_2, V_6, V_{11}, V_{12}, V_{17}, V_{24} pour la dimension "y". Les variables appartenant à une même dimension seront plus fortement corrélées. La matrice de corrélations est reproduite au tableau 6 : 30 boîtes ont été utilisées pour calculer ces corrélations (Thurstone [1947, p. 369]).

TABLEAU 6 : Matrice de corrélations de 12 variables du "Thurstone box problem"

	V1	V2	V9	V11	V5	V24	V6	V20	V10	V12	V15	V17
V1	100	24	89	24	76	29	26	72	61	26	59	-6
V2	24	100	33	89	37	80	79	35	35	67	-3	56
V9	89	33	100	47	95	54	54	93	87	58	21	-30
V11	24	89	47	100	59	94	96	58	61	90	-31	23
V5	76	37	95	59	100	68	69	99	97	75	-3	-40
V24	29	80	54	94	68	100	97	68	73	93	-37	5
V6	26	79	54	96	69	97	100	69	74	97	-42	2
V20	72	35	93	58	99	68	69	100	97	75	-8	-44
V10	61	35	87	61	97	73	74	97	100	83	-24	-50
V12	26	67	58	90	75	93	97	75	83	100	-50	-17
V15	59	-3	21	-31	-3	-37	-42	-8	-24	-50	100	43
V17	-6	56	-30	23	-40	5	2	-44	-50	-17	43	100

Dans le cas du type II, la matrice de corrélations comprend les mêmes variables que celle du type I sauf V_5 et V_{20}, nous assurant alors qu'il y a un nombre différent de variables (4 vs 6) par dimension.

La matrice de corrélations du type III provient de l'étude de Rimoldi, Bégin et alii [1982]. Les 12 variables regroupées en deux blocs (facteurs) de six variables chacun se lisent comme suit :

Tests intelligence :
- V_1 : Sommes
- V_2 : Synonymes
- V_5 : Divisions
- V_7 : Relations quantitatives
- V_9 : Matrices de figures
- V_{11} : Copies

Problèmes :
- V_{17} : Objets concrets 33
- V_{18} : Objets géométriques 31
- V_{19} : Objets géométriques 33
- V_{21} : Verbal qualitatif 33
- V_{23} : Verbal quantitatif 33
- V_{25} : Abstrait 33.

La matrice de corrélations de ces variablees (ou tests) a été calculée à partir de 180 sujets. On l'a reproduite au tableau 7.

C'est en enlevant les deux variables V_{23} et V_{25} de cette matrice que l'on a défini la matrice de corrélations du type IV. Les deux blocs ont alors respectivement 6 et 4 variables.

TABLEAU 7 : Matrice de corrélations de 12 tests psychologiques

	V1	V17	V7	V21	V9	V23	V11	V25	V2	V19	V5	V18
V1	100	6	71	9	54	11	53	16	52	14	67	3
V17	6	100	9	55	0	41	15	42	19	59	22	41
V7	71	9	100	11	59	17	40	8	46	13	47	1
V21	9	55	11	100	4	59	22	45	20	49	22	23
V9	54	0	59	4	100	3	36	-1	42	12	31	-1
V23	11	41	17	59	3	100	15	54	19	55	23	17
V11	53	15	40	22	36	15	100	22	51	22	60	21
V25	16	42	8	45	-1	54	22	100	30	35	33	19
V2	52	19	46	20	42	19	51	30	100	31	58	11
V19	14	59	13	49	12	55	22	35	31	100	29	24
V5	67	22	47	22	31	23	60	33	58	29	100	20
V18	3	41	1	23	1	17	21	19	13	27	20	100

Nous nous sommes inspirés du "Thurstone box problem" pour bâtir la matrice de corrélations du type V. Trente-six variables ont été regroupées autour des trois dimensions (facteurs) x, y et z :

V_1	:	x	,	V_{34}	:	e^y	,	V_{35} : e^z
V_{33}	:	e^x	,	V_2	:	y	,	V_3 : z
V_{30}	:	$\log x$	,	V_{31}	:	$\log y$	,	V_{32} : $\log z$
V_{70}	:	$\log(x^2+z)$	,	V_{72}	:	$\log(y^2+z)$	,	V_{74} : $\log(z^2+y)$
V_{69}	:	$\log(x^2+y)$	,	V_{71}	:	$\log(y^2+x)$	,	V_{73} : $\log(z^2+x)$
V_{54}	:	$x^2 \log y$	,	V_{51}	:	$y^4 z$	,	V_{53} : $z^4 y$
V_{48}	:	$x^4 y$	,	V_{50}	:	$y^4 x$	,	V_{47} : $z^3 y$
V_{42}	:	$x^3 y$	,	V_{45}	:	$y^3 z$	,	V_{59} : $z^2 \log y$
V_{49}	:	$x^4 z$	,	V_{56}	:	$y^2 \log x$	,	V_{52} : $z^4 x$
V_{55}	:	$x^2 \log z$	,	V_{44}	:	$y^3 x$	,	V_{46} : $z^3 x$
V_{43}	:	$x^3 z$	,	V_{57}	:	$y^2 \log z$	,	V_{12} : $z^2 y$
V_7	:	$x^2 y$	,	V_{11}	:	$y^2 z$	,	V_{58} : $z^2 \log x$

Un échantillon de 190 boîtes fictives, choisies à la manière de Thurstone [1947, p. 141], a permis de calculer la matrice de corrélations du tableau 8.

Pour le type VI, nous avons choisi la sous-matrice qui exclut les variables V_{13}, V_{14} et V_{24} de la matrice du type V.

TABLEAU 8 : Matrice de corrélations de 36 variables du "Thurstone box problem"

	V1	V33	V34	V2	V35	V3	V36	V70	V31	V72	V32	V74	V69	V54	V71	V61	V73	V53
V1	100	73	9	14	1	7	95	96	19	16	7	9	95	87	31	12	22	9
V33	73	100	7	6	0	0	69	63	7	6	8	8	61	72	15	13	17	11
V34	9	7	100	21	10	8	10	9	54	61	4	14	13	29	58	90	7	14
V2	14	6	21	100	10	11	16	14	95	96	10	26	32	43	95	77	12	41
V35	1	0	10	10	100	72	0	13	9	19	68	62	3	9	9	35	61	83
V3	7	9	8	11	72	100	5	23	9	26	95	96	7	12	10	42	96	95
V36	95	69	10	16	0	5	100	95	22	18	6	8	96	77	34	12	21	7
V70	96	63	9	14	13	23	95	100	20	21	24	29	94	81	32	17	41	19
V31	19	7	54	95	9	9	22	20	100	95	8	24	38	44	98	59	11	37
V72	16	6	61	96	19	26	18	21	95	100	25	42	34	43	95	66	21	44
V32	7	8	4	10	55	95	6	24	8	25	100	96	7	11	9	38	96	61
V74	9	8	16	26	61	96	8	29	24	42	96	100	12	17	25	45	96	68
V69	95	61	18	32	3	7	96	94	38	34	7	12	100	81	51	21	23	15
V54	87	72	29	43	9	12	77	81	44	43	11	17	81	100	50	35	54	24
V71	31	15	58	95	9	10	34	32	98	95	9	25	51	50	100	62	14	38
V61	12	12	79	73	35	42	12	17	59	66	38	45	21	35	62	100	40	70
V73	22	17	7	12	61	96	21	41	11	27	96	96	23	24	14	40	100	66
V53	9	11	14	41	83	95	6	19	37	44	61	68	15	24	38	70	66	100
V48	77	80	31	37	11	13	64	69	35	36	12	17	68	95	40	3	23	16
V42	80	75	34	42	10	13	68	73	40	41	12	18	72	97	45	40	23	16
V50	41	32	82	74	10	12	38	39	60	65	10	21	43	67	64	75	16	35
V45	12	12	74	75	39	48	12	19	63	70	44	51	22	36	66	99	45	75
V47	11	11	37	46	79	78	9	21	42	49	65	72	15	26	43	75	70	99
V59	14	11	31	49	72	81	13	25	49	54	72	78	21	29	49	70	76	95
V49	75	82	12	14	29	38	61	69	15	18	35	37	65	76	22	27	47	36
V55	81	73	11	16	27	45	71	77	18	21	46	46	74	78	26	36	51	56
V56	49	30	74	82	10	12	50	49	72	76	11	23	55	74	76	71	19	35
V44	46	35	77	76	10	13	43	45	64	69	11	22	49	74	68	73	19	35
V52	41	37	13	15	26	22	37	44	15	23	59	64	38	44	19	38	65	74
V46	46	40	13	16	71	74	42	49	16	25	62	67	43	48	21	39	66	74
V43	77	77	12	15	31	42	65	71	16	20	40	41	68	77	24	28	46	39
V7	81	67	39	50	10	13	72	75	48	49	13	19	76	98	53	44	23	27
V57	14	11	66	79	35	52	14	21	70	76	52	57	25	39	71	94	52	71
V11	13	11	66	76	45	56	13	21	67	74	52	59	24	37	69	96	54	66
V12	12	11	41	53	71	79	11	23	49	57	69	75	20	30	50	80	73	96
V58	49	35	13	18	66	77	49	54	18	27	69	73	49	47	24	41	74	71

	V48	V42	V50	V45	V47	V59	V49	V55	V56	V44	V52	V46	V43	V7	V57	V11	V12	V58
V1	77	80	41	12	11	14	75	81	49	46	41	46	77	81	14	13	12	49
V33	80	75	32	12	11	11	82	73	30	35	37	40	77	67	11	11	11	35
V34	31	34	82	74	37	31	12	11	74	77	13	13	12	39	66	66	41	13
V2	37	42	74	75	46	49	14	16	82	76	15	16	15	50	79	74	51	18
V35	11	10	10	39	79	72	29	27	10	10	76	71	31	10	35	45	71	56
V3	13	13	12	48	78	81	38	45	12	13	72	74	42	13	52	56	79	77
V36	64	68	38	12	9	13	61	71	50	43	37	42	65	72	14	13	11	49
V70	69	73	39	19	21	25	68	77	49	45	44	49	71	76	20	21	23	54
V31	35	40	60	63	42	49	15	18	72	64	15	16	16	48	20	67	69	18
V72	36	41	65	70	48	54	18	21	76	69	15	25	20	49	26	24	57	23
V32	12	12	10	44	65	72	35	46	11	11	59	62	40	12	52	52	69	69
V74	17	18	21	51	72	78	37	46	23	23	64	67	41	19	52	99	75	71
V69	68	72	43	22	16	21	65	74	55	49	38	43	68	76	25	24	20	49
V54	95	97	67	36	26	29	76	78	74	74	44	49	77	98	19	17	30	47
V71	40	45	64	66	43	49	22	26	76	68	19	21	24	53	71	69	59	24
V61	37	40	75	99	75	70	27	29	71	73	39	39	28	44	44	96	90	41
V73	22	23	16	46	70	76	42	51	19	18	65	68	46	23	52	54	73	74
V53	25	26	35	75	99	95	36	46	35	35	2	24	39	27	51	80	94	71
V48	100	99	70	38	27	27	79	76	69	74	43	47	78	96	38	38	30	43
V42	99	100	73	40	28	29	77	76	74	78	43	47	77	99	41	40	31	45
V50	70	73	100	73	39	36	40	41	95	99	28	30	41	78	70	68	45	30
V45	38	40	73	100	80	76	28	30	71	72	43	44	30	45	97	99	65	45
V47	27	28	39	80	100	97	37	38	40	40	75	73	41	30	77	85	99	72
V59	27	29	36	76	97	100	40	43	40	38	74	74	44	32	78	84	98	75
V49	79	77	40	28	37	40	100	95	40	44	23	27	99	73	29	31	38	71
V55	76	76	41	30	38	43	95	100	44	45	30	36	97	74	34	34	41	73
V56	69	74	95	71	40	40	40	44	100	97	30	33	42	82	72	69	47	35
V44	74	78	99	72	40	38	44	45	97	100	30	33	45	83	76	68	46	53
V52	43	43	28	43	75	74	23	30	30	30	100	99	26	43	41	49	71	64
V46	47	47	30	44	73	74	27	36	33	33	99	100	80	47	43	50	71	67
V43	78	77	41	30	41	44	99	97	42	45	26	80	100	74	31	34	41	76
V7	96	99	78	45	30	32	73	74	82	83	43	47	74	100	48	45	34	46
V57	38	41	70	97	77	78	29	34	72	70	41	43	31	46	100	99	95	67
V11	38	40	68	99	85	84	31	34	69	68	49	50	34	45	98	100	91	52
V12	30	31	45	65	99	98	38	41	47	46	71	71	41	34	95	91	100	71
V58	43	45	30	45	72	75	71	73	35	33	64	67	76	46	67	52	71	100

Afin de bâtir la matrice du type VII, 42 items (répondus par 750 élèves) ont été choisis à partir d'un inventaire d'intérêts de 190 items (Rhéaume [1982]). Ces 42 variables sont regroupées en deux échelles (facteurs) : une échelle d'ordre littéraire comprenant les variables dont les numéros sont 4, 18, 23, 35, 47, 55, 58, 65, 69, 75, 80, 86, 91, 93, 97, 106, 115, 122, 131, 149, 187 ; et une échelle d'ordre technique comportant les variables 1, 10, 28, 31, 38, 43, 50, 63, 72, 79, 84, 98, 119, 130, 140, 145, 146, 153, 170, 178, 189.

La matrice de corrélations de ces variables se trouve au tableau 9.

Il faut enlever les variables 31, 35, 130 et 189 pour obtenir la matrice correspondante du type VIII.

TABLEAU 9 : : Matrices de corrélations de 42 items d'un inventaire d'intérêts

	V84	V170	V97	V122	V189	V130	V63	V119	V92	V65	V91	V115	V72	V145	V10	V28	V58	V106	V4	V86	V38
V84	100	65	-3	-6	57	57	53	54	-2	-1	-5	1	46	46	50	51	-1	1	-1	-4	44
V170	65	100	-11	-8	71	72	48	69	-6	0	-9	-6	54	44	51	57	-5	-7	-16	-11	39
V97	-3	-11	100	46	-5	-7	0	-3	50	40	43	52	-1	5	-6	-7	1	45	50	47	5
V122	-6	-8	46	100	-3	-7	-1	-3	52	41	44	48	1	3	-12	-9	37	41	36	37	9
V189	57	71	-5	-3	100	63	40	58	-5	9	1	-3	49	39	44	48	-1	0	-11	-6	35
V130	57	72	-7	-7	63	100	41	71	-3	9	-4	-5	62	55	47	49	0	5	-14	-5	35
V63	53	48	0	-1	40	41	100	38	-1	6	2	6	38	54	47	46	4	3	-9	1	4
V119	54	69	-3	-3	58	71	38	100	-2	8	-9	-4	56	27	43	43	3	5	-11	-3	21
V92	-2	-6	50	52	-5	-3	-1	-2	100	40	32	61	-1	4	-14	-8	34	47	52	36	8
V65	-1	0	40	41	5	9	6	8	40	100	38	31	15	1	-5	-2	36	36	32	39	3
V91	-5	-9	43	46	1	-4	2	-9	32	38	100	30	-2	5	-12	-11	37	36	30	33	0
V115	1	-6	52	48	-3	-6	6	-4	61	31	30	100	-1	8	-6	-6	33	41	48	23	5
V72	46	54	-1	1	49	62	36	56	-1	12	2	-1	100	32	43	43	8	9	-8	-3	33
V145	46	44	6	2	39	55	54	27	4	1	5	8	32	100	44	46	3	8	-3	6	39
V10	50	51	-6	-12	44	47	47	43	-14	5	-12	-6	43	44	100	53	-4	-1	-12	-9	35
V28	51	57	-7	-8	48	49	46	43	-8	-6	-11	-6	43	40	53	100	-6	0	-11	-12	32
V58	-1	-6	47	37	-1	0	4	3	34	36	37	33	8	3	-4	-6	100	40	34	59	2
V106	1	-3	45	41	0	5	3	-5	47	36	36	41	9	8	-1	0	40	100	37	35	11
V4	-1	-16	50	36	-12	-14	-8	-11	52	32	30	48	-8	-3	-12	-11	34	37	100	35	0
V86	-4	-11	42	37	-8	-5	1	-3	36	38	33	33	-3	0	-9	-12	59	35	35	100	4
V38	44	38	-2	0	30	36	42	31	8	3	0	5	33	36	35	32	2	11	0	4	100
V98	42	36	6	3	40	29	41	35	4	1	14	9	26	49	32	29	5	2	-3	5	35
V146	34	36	-5	-1	36	30	39	21	-4	5	-1	0	39	37	34	30	-4	-5	-13	-9	33
V178	45	49	2	8	52	46	29	47	8	10	10	6	38	26	34	30	9	13	-1	1	35
V69	-2	1	40	52	3	9	5	2	51	34	31	42	6	9	-2	-2	28	44	41	29	9
V131	-5	-11	40	34	-7	-9	-7	-5	31	14	44	26	-4	-4	-12	-10	35	14	38	30	-4
V55	-12	-18	44	39	-11	-16	-10	-10	31	33	40	23	-9	-5	-19	-16	34	32	40	30	-9
V23	1	-11	36	30	-5	-6	1	-3	21	31	34	18	-3	6	-4	-6	32	34	30	26	1
V50	41	46	-11	-7	45	37	44	30	-7	-7	-1	-7	30	44	40	45	-17	-3	-14	-17	30
V153	32	36	-5	-7	35	34	37	28	-4	5	-2	1	27	36	28	34	0	-2	-15	-1	40
V79	41	37	14	3	30	41	23	47	7	7	-1	3	39	22	40	35	11	11	3	6	33
V140	31	31	6	11	30	38	24	38	11	7	0	7	42	25	25	24	6	16	0	5	30
V80	-2	-2	37	32	-2	0	-2	3	54	56	30	25	-1	0	-10	-5	21	26	24	25	-1
V75	-5	-11	40	27	-6	-7	-6	-5	27	29	47	24	-8	-3	-9	-13	32	31	39	27	-3
V18	-2	-8	33	33	-2	0	4	-1	23	29	29	27	4	7	-1	-1	14	29	30	32	1
V149	-2	-2	32	33	1	6	-1	3	40	32	22	36	12	4	-3	-2	30	32	31	30	1
V1	36	35	4	3	35	38	21	39	2	10	4	-1	32	17	31	23	10	9	0	9	21
V43	33	19	8	6	17	19	31	14	5	4	8	2	24	33	16	25	10	12	9	5	33
V31	29	18	6	7	18	16	48	14	10	11	14	9	18	42	26	32	14	9	1	8	29
V187	-2	-4	33	34	9	5	-5	3	26	42	34	29	12	5	-7	-5	26	28	26	30	-3
V47	-4	-3	28	25	-2	3	1	4	21	41	20	21	-1	-1	-7	-5	33	33	33	26	-1
V35	10	1	34	16	4	6	15	7	17	27	29	21	2	11	10	1	32	31	24	26	13

	V98	V146	V178	V69	V131	V55	V23	V50	V153	V79	V140	V80	V75	V18	V149	V1	V43	V31	V187	V47	V35
V84	42	34	45	2	-5	-12	1	41	32	41	31	-2	-9	-5	-2	36	33	29	-2	-4	10
V170	36	36	49	1	-11	-18	-11	46	36	37	31	-2	-11	-8	-2	35	18	18	-4	-3	1
V97	6	-5	2	40	40	44	36	-11	-5	14	6	37	40	37	32	4	8	6	33	28	34
V122	3	-1	8	52	34	39	30	-2	7	3	11	32	27	23	33	3	6	7	34	25	16
V189	40	36	52	3	-7	-11	-5	45	35	30	30	3	-6	2	1	32	17	18	9	-3	4
V130	29	30	46	9	-9	-16	-6	37	34	41	38	0	-7	0	6	38	19	16	5	1	6
V63	41	39	29	5	-7	-10	1	44	37	23	24	5	-6	4	-1	21	31	48	-5	1	15
V119	20	21	47	7	-5	-10	-3	30	28	47	38	3	-5	-1	3	39	14	14	3	4	7
V93	4	-4	8	51	31	31	21	-7	-4	7	11	29	27	23	40	2	5	10	26	21	17
V65	1	2	10	34	34	33	31	-7	5	7	7	56	29	29	32	10	4	11	42	41	27
V91	14	-1	10	31	44	48	34	-1	7	-1	0	30	47	39	35	4	11	14	34	29	28
V115	9	0	6	42	25	23	18	-7	1	3	7	25	24	27	36	-1	2	9	29	21	21
V72	26	29	38	5	-4	-9	-3	30	27	39	42	-1	-8	4	12	35	24	18	12	-1	2
V145	49	37	26	9	-4	-5	6	44	36	32	25	0	-3	7	4	17	33	42	5	-1	11
V10	32	34	34	-2	-12	-19	-4	40	28	40	25	-10	-9	-1	-3	31	30	26	-7	-7	10
V28	29	30	30	-2	-10	-16	-6	45	34	25	24	-5	-13	-1	2	23	25	22	-5	-5	1
V58	5	-4	9	28	35	34	32	-13	0	11	6	21	32	34	30	10	10	14	26	23	32
V106	7	-5	13	44	34	32	34	-3	2	11	16	26	31	29	32	9	12	9	28	22	33
V4	-2	-13	-1	41	38	40	30	-14	-15	5	0	24	39	30	31	0	9	4	26	22	24
V86	5	-8	1	29	30	30	26	-12	-1	6	5	25	27	32	30	6	5	8	30	20	26
V38	35	22	28	9	-4	-6	1	30	22	31	38	-1	-3	1	1	21	33	29	-3	-1	13
V98	100	43	27	5	-2	-6	10	37	36	17	19	8	2	5	1	19	24	34	8	0	10
V146	43	100	28	-2	-9	-16	-4	46	49	8	16	0	-5	-1	1	16	20	31	3	-6	6
V178	27	28	100	13	6	3	9	24	22	30	24	6	4	5	2	31	16	16	7	7	14
V69	5	-2	13	100	21	20	17	-4	6	11	19	26	19	26	47	3	5	7	34	23	16
V131	-2	-9	6	21	100	61	55	-16	-10	7	1	20	61	16	13	7	10	6	17	16	35
V55	-6	-16	3	20	61	100	54	-20	-15	6	-4	27	52	27	10	2	5	2	26	23	34
V23	10	-4	9	17	55	54	100	-4	-7	9	-2	23	49	23	10	10	9	16	18	18	38
V50	37	46	24	-4	-16	-20	-4	100	40	12	13	-9	-10	-8	-6	16	19	25	0	-13	5
V153	36	49	22	6	-10	-15	-7	40	100	9	24	3	-8	8	3	18	12	26	12	5	6
V79	17	8	30	11	7	6	9	12	9	100	37	6	9	6	9	37	28	11	4	6	12
V140	19	16	24	19	1	-4	-2	13	24	37	100	5	-3	4	14	22	21	8	6	5	4
V80	8	0	6	26	20	27	23	-7	3	8	5	100	22	32	24	8	1	7	29	53	19
V75	2	-5	4	19	61	52	49	-10	-8	9	-1	22	100	19	14	8	11	2	15	15	31
V18	5	-1	5	26	16	27	23	-8	8	6	4	32	19	100	25	5	10	12	37	32	24
V149	1	1	2	43	13	18	10	-6	3	9	14	24	14	25	100	7	2	-1	41	19	14
V1	19	16	31	3	7	2	10	16	18	37	22	8	8	5	7	100	14	9	8	6	14
V43	24	20	16	5	10	5	9	19	12	28	21	1	11	10	2	14	100	31	8	1	10
V31	34	31	16	7	6	2	16	25	26	11	8	7	7	12	-1	9	31	100	7	8	13
V187	8	3	7	34	17	26	18	0	12	4	6	29	15	37	41	8	8	7	100	21	11
V47	0	-6	7	23	16	23	18	-13	5	6	5	53	15	32	19	6	1	8	21	100	22
V35	10	6	14	16	35	34	38	5	6	12	4	19	31	24	14	14	10	13	11	22	100

II.2.2 Les méthodes d'extraction et de rotation de l'étude.

Une fois construites ces huit matrices de corrélations, nous avons obtenu pour chaque matrice, 45 solutions différentes. C'est-à-dire que nous avons utilisé cinq méthodes d'extraction combinées à neuf méthodes de rotation pour chaque matrice. Les cinq méthodes d'extraction employées sont 1) l'analyse en composantes principales (A.C.P.), 2) la méthode des facteurs principaux avec estimé de communauté, la valeur absolue du plus grand coefficient de corrélation pour chaque variable (A.F.P. (Maxrow)), 3) la méthode des facteurs principaux avec estimé de communauté, le coefficient de corrélation multiple au carré (A.F.P. (SMC)), 4) la méthode alpha avec estimé de communauté, le coefficient de corrélation multiple au carré (ALPHA (SMC)) 5) la méthode du maximum de vraisemblance avec estimé de communauté, le coefficient de corrélation multiple au carré (MALIFA (SMC)). Les neuf méthodes de rotation employées sont

1) orthomax ($\gamma = 0$) i.e. quartimax
2) orthomax ($\gamma = 0,25$)
3) orthomax ($\gamma = 0,50$) i.e. biquartimax
4) orthomax ($\gamma = 0,75$)
5) orthomax ($\gamma = 1$) i.e. varimax
6) promax
7) orthoblique
8) biquartimin
9) maxplane.

Toutes ces méthodes d'extraction et de rotation ont été décrites succintement au cours de la première partie de ce texte.

II.2.3 La liste des critères

Une fois les 45 solutions trouvées pour chaque matrice de corrélations, l'objectif est de trouver les solutions les plus adéquates ou acceptables compte tenu du type retenu ; il s'agit aussi bien sûr de révéler les solutions les moins adéquates pour chaque matrice de corrélations.

Afin de juger de l'adéquation de l'une ou l'autre des solutions, une liste de quatre critères a été établie :

1) la distance par rapport à une "solution de compromis" construite à l'aide des 45 solutions,

2) le temps informatique nécessaire pour produire les facteurs avant et après rotation,

3) l'invariance factorielle au sens de Kaiser [1958] tel qu'interprétée par Bertrand et Dufour [1981],

3) le choix personnel de spécialistes reconnus.

Les quatre sections suivantes décrivent plus amplement ces quatre critères ainsi que l'utilisation qui en est faite pour établir l'adéquation (ou l'inadéquation !) de certaines solutions.

Le tableau 10 présente un système de codage double des 45 méthodes (extraction-rotation) utilisées. Ce codage nous permettra de mieux traiter de la comparaison entre les solutions correspondantes.

Tableau 10 : Codage des méthodes (extraction-rotation) utilisées.

A =	1	A.C.P. - orthomax ($\gamma = 0$)
	2	A.C.P. - orthomax ($\gamma = 0,25$)
	3	A.C.P. - orthomax ($\gamma = 0,50$)
	4	A.C.P. - orthomax ($\gamma = 0,75$)
	5	A.C.P. - orthomax ($\gamma = 1$)
B =	6	A.C.P. - promax
C =	7	A.C.P. - orthoblique
D =	8	A.C.P. - biquartimin
E =	9	A.C.P. - maxplane
F =	10	A.F.P. (MAXROW) - orthomax ($\gamma = 0$)
	11	A.F.P. (MAXROW) - orthomax ($\gamma = 0,25$)
	12	A.F.P. (MAXROW) - orthomax ($\gamma = 0,50$)
	13	A.F.P. (MAXROW) - orthomax ($\gamma = 0,75$)
	14	A.F.P. (MAXROW) - orthomax ($\gamma = 1$)
G =	15	A.F.P. (MAXROW) - promax
H =	16	A.F.P. (MAXROW) - orthoblique
I =	17	A.F.P. (MAXROW) - biquartimin
J =	18	A.F.P. (MAXROW) - maxplane
K =	19	A.F.P. (SMC) - orthomax ($\gamma = 0$)
	20	A.F.P. (SMC) - orthomax ($\gamma = 0,25$)
	21	A.F.P. (SMC) - orthomax ($\gamma = 0,50$)
	22	A.F.P. (SMC) - orthomax ($\gamma = 0,75$)
	23	A.F.P. (SMC) - orthomax ($\gamma = 1$)

L = 24 : A.F.P. (SMC) - promax
M = 25 : A.F.P. (SMC) - orthoblique
N = 26 : A.F.P. (SMC) - biquartimin
O = 27 : A.F.P. (SMC) - maxplane

P =
- 28 : Alpha (SMC) - orthomax (γ = 0)
- 29 : Alpha (SMC) - orthomax (γ = 0,25)
- 30 : Alpha (SMC) - orthomax (γ = 0,50)
- 31 : Alpha (SMC) - orthomax (γ = 0,75)
- 32 : Alpha (SMC) - orthomax (γ = 1)

Q = 33 : Alpha (SMC) - promax
R = 34 : Alpha (SMC) - orthoblique
S = 35 : Alpha (SMC) - biquartimin
T = 36 : Alpha (SMC) - maxplane

U =
- 37 : Malifa (SMC) - orthomax (γ = 0)
- 38 : Malifa (SMC) - orthomax (γ = 0,25)
- 39 : Malifa (SMC) - orthomax (γ = 0,50)
- 40 : Malifa (SMC) - orthomax (γ = 0,75)
- 41 : Malifa (SMC) - orthomax (γ = 1)

V = 42 : Malifa (SMC) - promax
W = 43 : Malifa (SMC) - orthoblique
X = 44 : Malifa (SMC) - biquartimin
Y = 45 : Malifa (SMC) - maxplane

Ainsi lorsque l'on fera référence à la solution A, cela signifiera une solution provenant de la méthode d'extraction, l'analyse en composantes principales, et d'une des méthodes de rotation orthomax. Par ailleurs, la solution 13 sera celle qui est issue de la méthode d'extraction, l'analyse en facteurs principaux où l'estimé des communautés est la valeur absolue du plus grand coefficient de corrélation pour chaque variable, et de la méthode de rotation orthomax où γ est fixé à 0,75. Dépendant du contexte, il pourra être plus utile d'utiliser soit le codage alphabétique soit le codage numérique.

II.3 Critère 1 : distance par rapport à la solution de compromis.

II.3.1 Description

Ce critère permettra de mieux détecter les solutions dites aberrantes, c'est-à-dire celles qui sont isolées des autres, celles dont le pattern est relativement différent des autres solutions.

Afin d'utiliser ce critère de façon efficiente, nous avons construit, pour chacune des huit matrices de corrélations, une solution de compromis de la façon suivante. Nous avons d'abord obtenu les 45 patterns factoriels. Puis on a identifié, pour chaque variable et chaque facteur, la saturation médiane[1] parmi l'ensemble des 45 saturations, créant ainsi un nouveau pattern dit médian ou de compromis.

---

1 La saturation médiane est définie comme celle qui se trouve à la 23e = $(\frac{1 + 45}{2})^e$ position lorsque les 45 saturations sont ordonnées.

On a par la suite calculé la distance facteur à facteur entre chacun des 45 patterns et ce pattern médian à l'aide de la formule :

$$d_i = \sum_{j=1}^{n} \frac{(a_{ji} - \bar{a}_{ji})^2}{4n} \qquad \text{pour le facteur Fi} \quad 1 \leq i \leq m$$

où n est le nombre de variables,

a_{ji} , les saturations du pattern d'une des 45 solutions,

$\bar{a}_{ji}$, les saturations du pattern médian

Comme cette distance est calculée facteur à facteur, il y aura donc autant d'indices d_i que de facteurs dans le pattern factoriel. Par exemple s'il y a deux facteurs F_1, F_2 pour chacun des 45 patterns, alors le couple (d_1, d_2) sera un indice de l'écart entre un des 45 patterns et le pattern médian.

La distance minimale nous est donnée dans le cas où $a_{ji} = \bar{a}_{ji}$ pour tous les j ; alors $d_i = 0$.

On obtient la distance maximale lorsque l'écart entre chaque a_{ji} et chaque $\bar{a}_{ji}$ est 2 (ex. $a_{ji} = -1$, $\bar{a}_{ji} = 1$).

d'où $$d_i = \sum_{j=1}^{n} \frac{2^2}{4n} = \frac{4n}{4n} = 1.$$

Ainsi $0 \leq d_i \leq 1$.

Incidemment notre distance d_i est reliée à l'indice r_{msi} connue sous le nom de "root mean square" (voir Harman [1976, p. 297]) par la formule :

$$d_i = \frac{r^2 msi}{4}$$

L'avantage de d_i est qu'il varie à l'intérieur de limites plus conventionnelles, 0 et 1.

Nous dirons qu'une solution est aberrante si au moins un des d_i est plus grand que le seuil 0,0025. Comme $\sum_{j=1}^{n} \frac{(0,1)^2}{4n} = 0,0025$, ce seuil correspond à un écart moyen entre les saturations a_{ji} et $\bar{a}_{ji}$ de 0,1 i.e. $| a_{ji} - \bar{a}_{ji} | = 0,1$.

Nous soulignerons tout particulièrement les solutions aberrantes pour lesquelles un des di est supérieur à 0,01 : ce qui correspond à un écart moyen $| a_{ji} - \bar{a}_{ji} |$ d'au moins 0,2.

II.3.2 Le type I

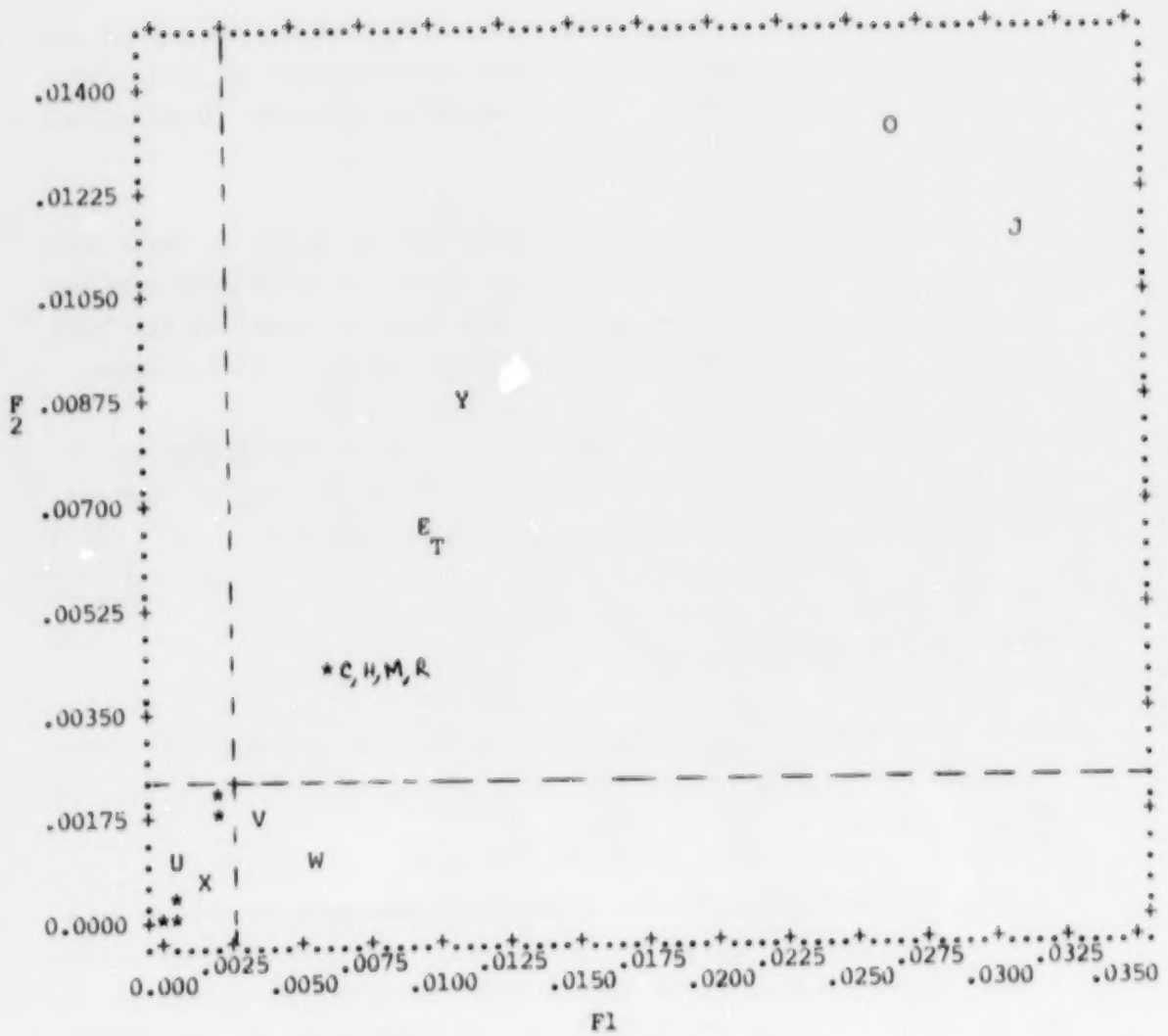
Le tableau 11 présente les distances d_i, facteur à facteur, de chacun des 45 patterns par rapport au pattern médian. Comme on peut le vérifier, onze solutions comportent un d_i plus grand que 0,0025 et sont donc considérées comme aberrantes. Il s'agit des solutions codées C, E, H, J, M, O, R, T, V, W, Y.

Le graphe 7, construit à partir du tableau 11, permet de détecter ces solutions aberrantes. Une ligne pointillée horizontale a été

TABLEAU 11 : Distance des 45 solutions par rapport à la solution médiane : type I

	FACTEUR 1	FACTEUR 2
1	.0004434	.0000149
2	.0003396	.0000209
3	.0002702	.0000388
4	.0002221	.0000612
5	.0001876	.0000850
6	.0021243	.0019354
7	.0060329	.0043016
8	.0002732	.0004144
9	.0096373	.0067511
10	.0004453	.0000147
11	.0003383	.0000215
12	.0002672	.0000405
13	.0002182	.0000643
14	.0001833	.0000892
15	.0020991	.0018982
16	.0060473	.0042468
17	.0002675	.0004030
18	.0307957	.0116902
19	.0004511	.0000154
20	.0003432	.0000214
21	.0002715	.0000399
22	.0002220	.0000633
23	.0001868	.0000879
24	.0020942	.0018996
25	.0060407	.0042541
26	.0002661	.0004078
27	.0262973	.0133169
28	.0004789	.0000210
29	.0003640	.0000230
30	.0002877	.0000391
31	.0002351	.0000611
32	.0001977	.0000849
33	.0020738	.0019412
34	.0059646	.0043172
35	.0002552	.0004299
36	.0097773	.0063579
37	.0002637	.0009700
38	.0002630	.0010217
39	.0002652	.0010644
40	.0002686	.0011001
41	.0002725	.0011302
42	.0036401	.0017162
43	.0053643	.0010136
44	.0012619	.0007142
45	.0111708	.0088044

Graphe 7 : Représentation cartésienne des 45 solutions : type I



tracée au niveau du seuil 0,0025 pour le facteur 2 : toute solution se situant en haut de cette ligne est donc considérée aberrante. Une ligne pointillée verticale a été tracée au niveau du seuil 0,0025 pour le facteur 1 : toute solution se situant à droite de cette ligne est aussi considérée aberrante.

En utilisant le code connu (tableau 10), on se rend compte que toutes ces solutions aberrantes sauf une, correspondent aux cinq orthoblique et aux cinq maxplane. L'autre solution aberrante (à peine) est une malifa (SMC) - promax.

Qui plus est, trois solutions méritent une attention toute particulière : J, O, Y. Leur distance par rapport à la solution médiane, pour l'un ou l'autre facteur, dépasse même 0,01. À noter que ces trois solutions ont été obtenues à partir de la méthode de rotation maxplane.

Remarquons aussi qu'en général les d_1 sont plus grands que les d_2 : c'est-à-dire les distances par rapport à la solution de compromis sont plus grandes pour le facteur 1 que pour le facteur 2.

II.3.3 Le type II

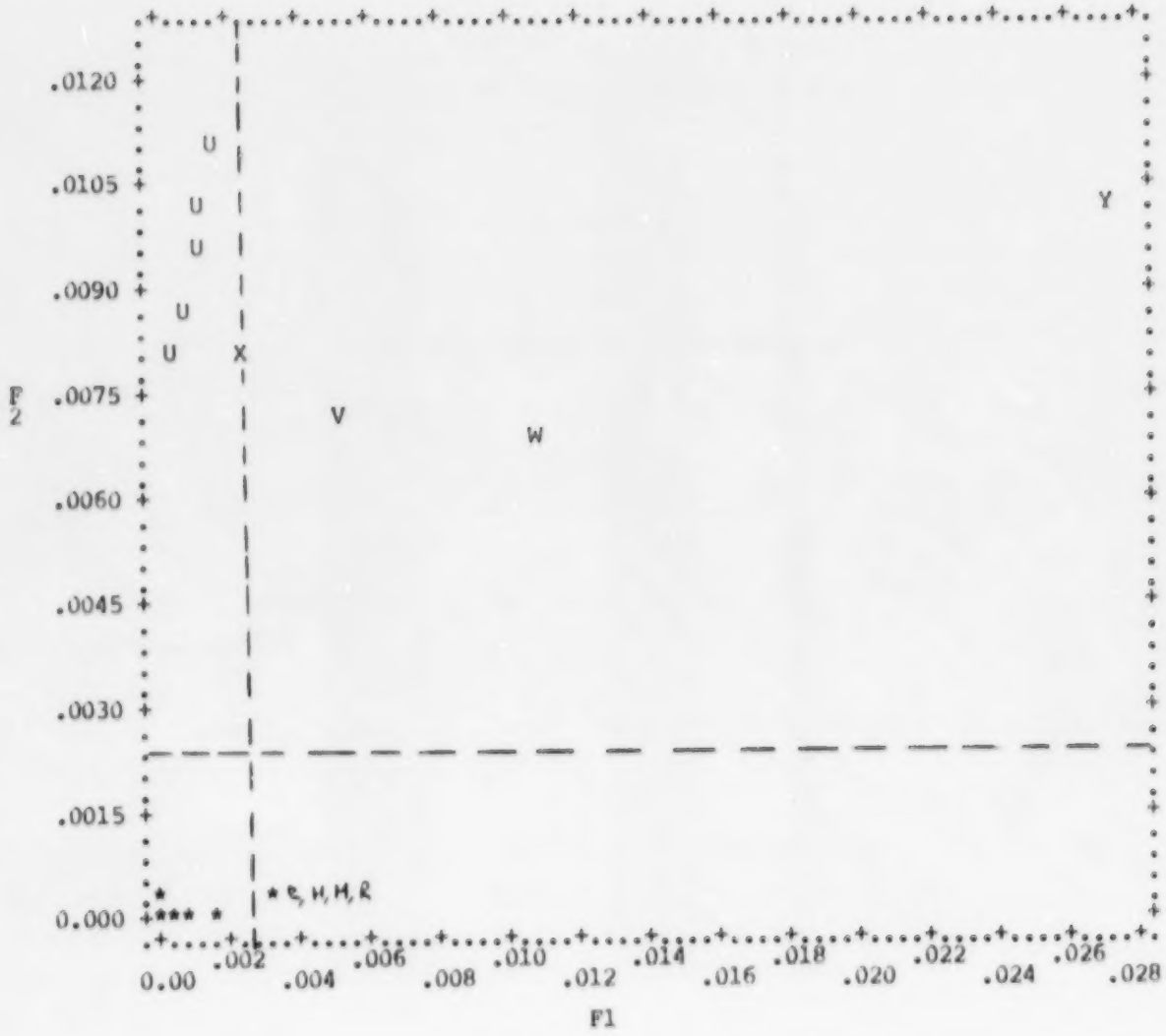
Le tableau 12 et le graphe 8 permettent de constater que treize solutions peuvent être considérées comme aberrantes : C, H, M, R, U (5 fois), V, W, X, Y.

On y retrouve toutes les solutions issues de la méthode de rotation orthoblique et toutes celles de la méthode d'extraction malifa (SMC). Quatre solutions sont spécialement éloignées des autres (d_1 ou $d_2 \geq 0{,}01$) : elles proviennent de la combinaison de la méthode

TABLEAU 12 : Distances des 45 solutions par rapport à la solution médiane : type II

	FACTEUR 1	FACTEUR 2
1	.0003334	.0000319
2	.0002438	.0000096
3	.0001640	.0000229
4	.0000971	.0000786
5	.0000466	.0001838
6	.0015047	.0000369
7	.0032790	.0003582
8	.0000124	.0000409
9	.0004558	.0001093
10	.0003108	.0000217
11	.0002226	.0000083
12	.0001448	.0000322
13	.0000805	.0001005
14	.0000336	.0002207
15	.0015644	.0000196
16	.0033986	.0003137
17	.0000057	.0000250
18	.0008486	.0001044
19	.0003107	.0000215
20	.0002226	.0000082
21	.0001448	.0000321
22	.0000806	.0001004
23	.0000337	.0002206
24	.0015643	.0000190
25	.0033979	.0003133
26	.0000059	.0000248
27	.0008491	.0001038
28	.0003919	.0000832
29	.0002943	.0000428
30	.0002056	.0000367
31	.0001289	.0000718
32	.0000677	.0001557
33	.0014200	.0000602
34	.0030675	.0003630
35	.0000111	.0000934
36	.0004727	.0000998
37	.0005985	.0082338
38	.0008003	.0088332
39	.0010457	.0095128
40	.0013357	.0102697
41	.0016697	.0110976
42	.0052312	.0072091
43	.0106853	.0069486
44	.0025879	.0081090
45	.0273130	.0101167

Graphe 8 : Représentation cartésienne des 45 solutions : type II



d'extraction malifa et des méthodes de rotation : orthomax ($\gamma = 0,75$), orthomax ($\gamma = 1$) ou varimax, orthoblique et maxplane. La combinaison malifa (SMC) - maxplane donne un pattern particulièrement distant du pattern médian.

Dans ce dernier cas en effet $d_1 = 0,027$ et $d_2 = 0,010$.

Tout comme pour le type I, même si c'est moins évident ici, les distances par rapport à la solution de compromis sont plus grandes pour le facteur 1 que pour le facteur 2.

II.3.4 Le type III

Comme on peut le voir au tableau 13 et au graphe 9, il n'existe aucune solution aberrante.

II.3.5 Le type IV

Ici encore aucune solution aberrante n'a été détectée : voir tableau 14 et graphe 10.

II.3.6 Le type V

Le tableau 15 nous montre que dix solutions sont aberrantes. Il s'agit des solutions B, C, G, H, L, M, Q, R, W, Y. Le graphe 11 donne une représentation visuelle des distances entre les solutions pour les facteurs 1 et 2, puis pour les facteurs 1 et 3 et enfin pour les facteurs 2 et 3.

TABLEAU 13 : Distance des 45 solutions par rapport à la solution médiane : type III

	FACTEUR 1	FACTEUR 2
1	.0004289	.0005491
2	.0004346	.0005467
3	.0004406	.0005450
4	.0004466	.0005437
5	.0004528	.0005430
6	.0015854	.0023700
7	.0011363	.0024398
8	.0005842	.0009326
9	.0014076	.0006681
10	.0000272	.0000464
11	.0000315	.0000402
12	.0000363	.0000354
13	.0000414	.0000315
14	.0000467	.0000286
15	.0012832	.0017831
16	.0006353	.0017909
17	.0002181	.0003400
18	.0003244	.0002795
19	.0000271	.0000464
20	.0000315	.0000402
21	.0000362	.0000353
22	.0000413	.0000315
23	.0000467	.0000286
24	.0012826	.0017830
25	.0006352	.0017909
26	.0002181	.0003399
27	.0003649	.0003717
28	.0000322	.0001375
29	.0000352	.0001291
30	.0000391	.0001224
31	.0000437	.0001171
32	.0000487	.0001131
33	.0013351	.0020424
34	.0006970	.0019285
35	.0002676	.0004141
36	.0006082	.0006410
37	.0001125	.0001008
38	.0001153	.0000900
39	.0001191	.0000813
40	.0001238	.0000745
41	.0001292	.0000692
42	.0012555	.0016117
43	.0002166	.0006622
44	.0003341	.0003130
45	.0003085	.0002493

Graphe 9 : Représentation cartésienne des 45 solutions : type III

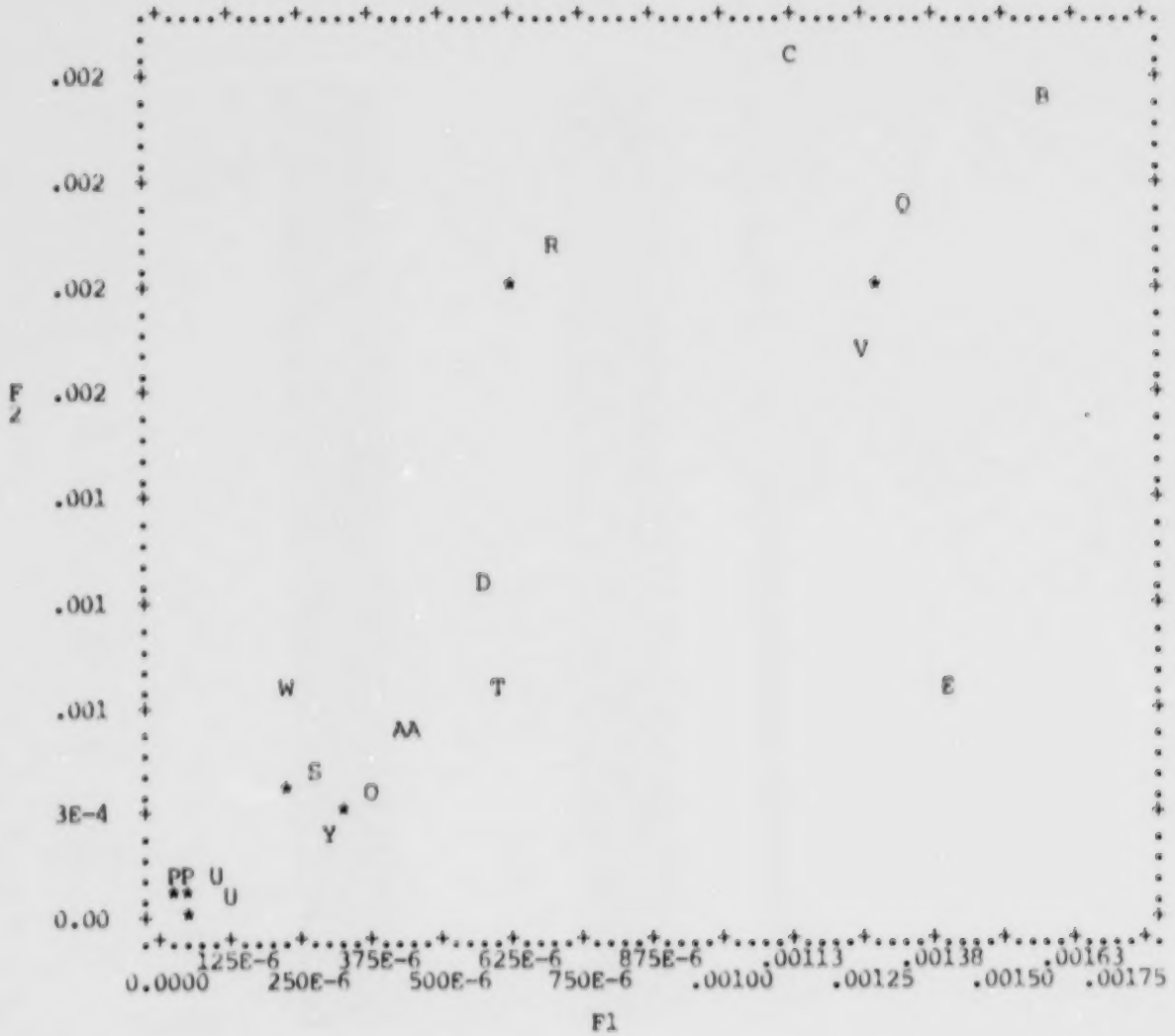


TABLEAU 14 : Distance des 45 solutions par rapport à la solution médiane : type IV

	FACTEUR 1	FACTEUR 2
1	.0006112	.0010146
2	.0005821	.0010209
3	.0005579	.0010337
4	.0005386	.0010526
5	.0005237	.0010771
6	.0017523	.0022477
7	.0011108	.0023589
8	.0006639	.0012449
9	.0014021	.0011983
10	.0000824	.0000362
11	.0000531	.0000503
12	.0000318	.0000749
13	.0000177	.0001082
14	.0000099	.0001492
15	.0014758	.0012619
16	.0005509	.0011746
17	.0001985	.0001889
18	.0008491	.0003310
19	.0000824	.0000361
20	.0000531	.0000503
21	.0000318	.0000748
22	.0000177	.0001082
23	.0000099	.0001491
24	.0014756	.0012627
25	.0005510	.0011745
26	.0001985	.0001890
27	.0005079	.0009476
28	.0000568	.0001592
29	.0000378	.0001856
30	.0000267	.0002214
31	.0000224	.0002649
32	.0000240	.0003146
33	.0017006	.0016133
34	.0008675	.0016551
35	.0003230	.0002612
36	.0006945	.0003209
37	.0001367	.0001014
38	.0001067	.0001077
39	.0000842	.0001239
40	.0000684	.0001489
41	.0000588	.0001814
42	.0016604	.0011672
43	.0001885	.0005587
44	.0002161	.0002934
45	.0003844	.0012363

Graphe 10 : Représentation cartésienne des 45 solutions : type IV

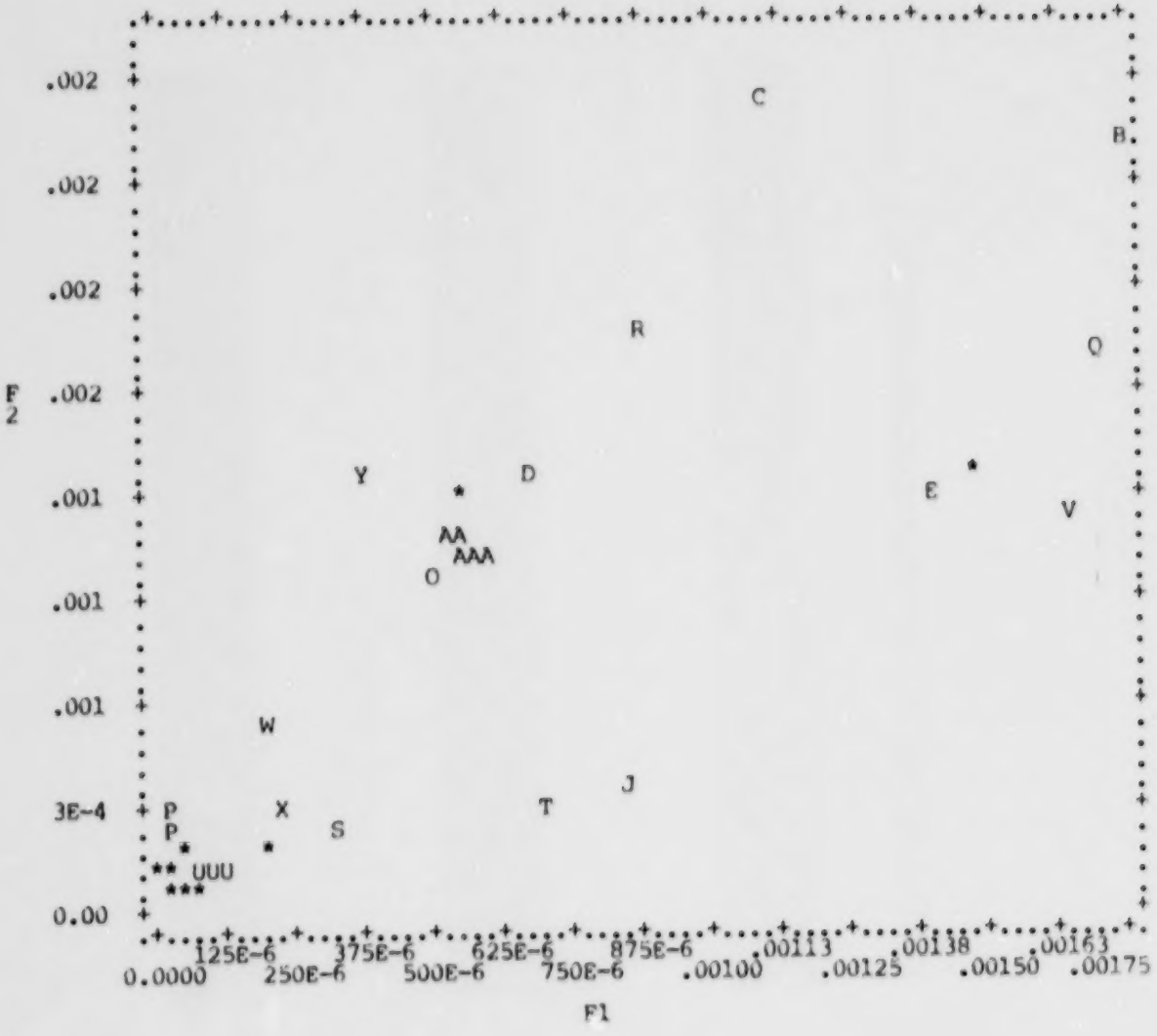
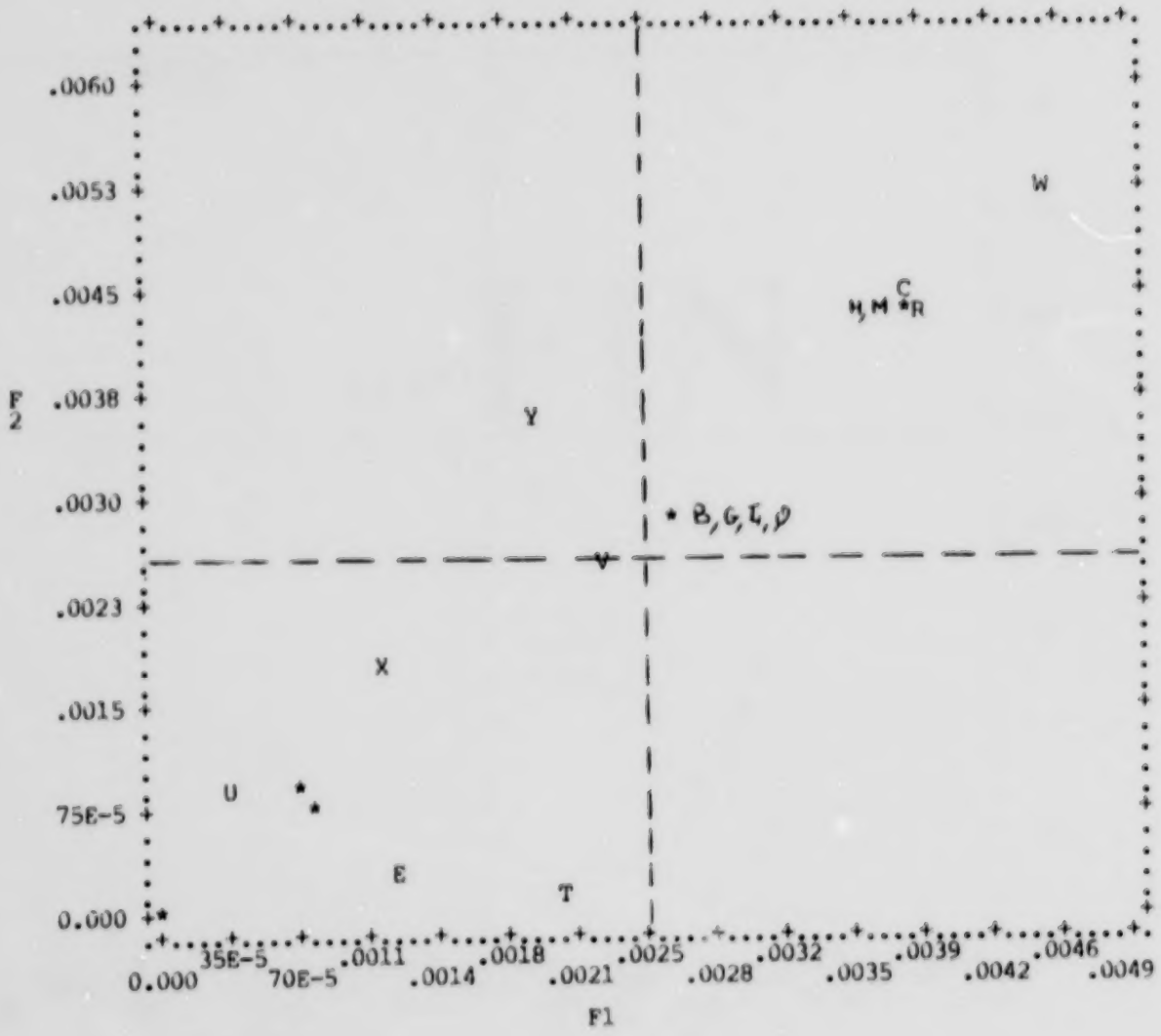


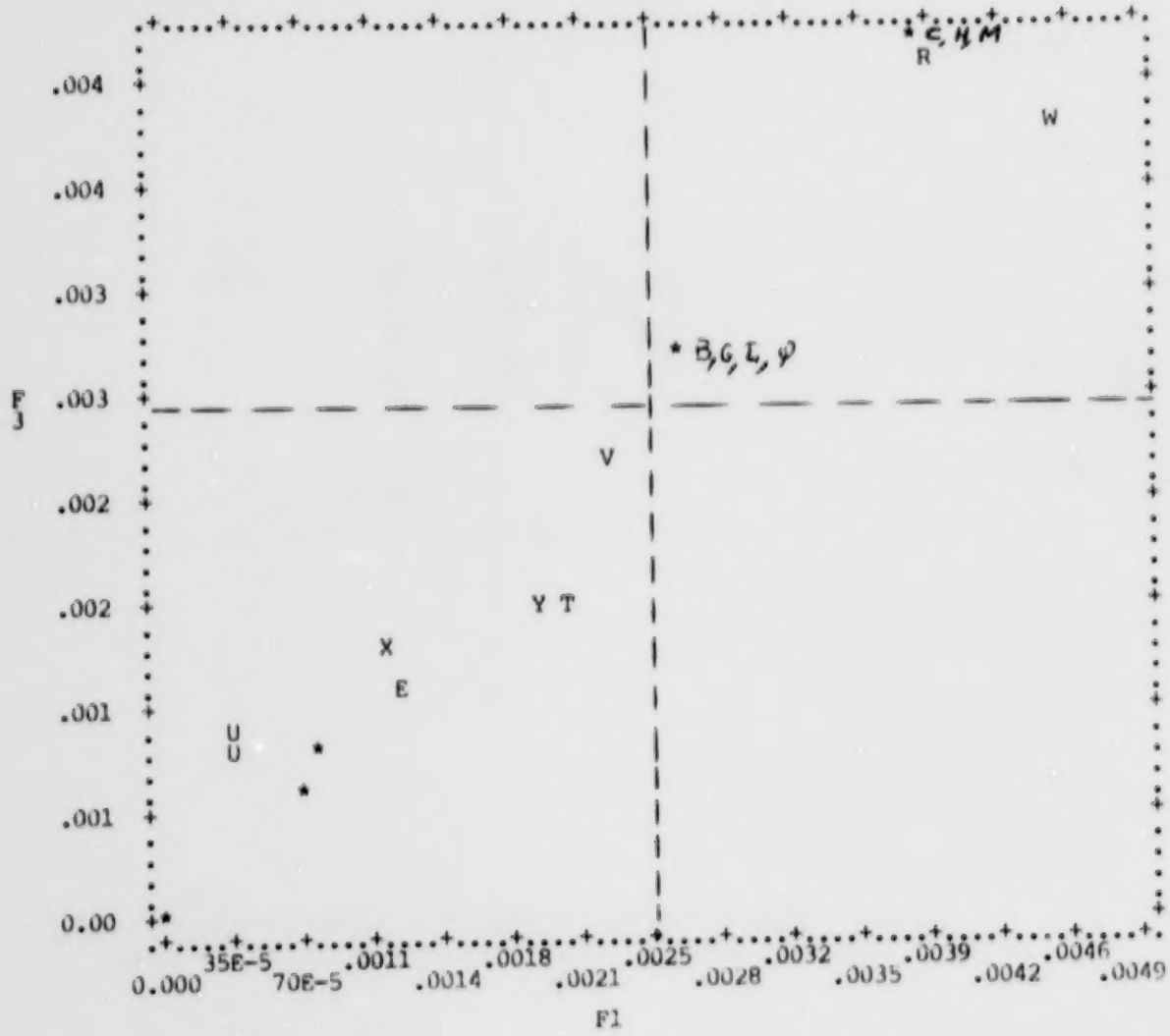
TABLEAU 15 : Distance des 45 solutions par rapport à la solution médiane : type V

	FACTEUR 1	FACTEUR 2	FACTEUR 3
1	.0000178	.0000355	.0000265
2	.0000147	.0000403	.0000255
3	.0000124	.0000453	.0000249
4	.0000106	.0000503	.0000247
5	.0000093	.0000552	.0000246
6	.0026225	.0028371	.0026995
7	.0038032	.0044564	.0041714
8	.0007524	.0007591	.0007732
9	.0012032	.0002690	.0011465
10	.0000218	.0000328	.0000218
11	.0000181	.0000379	.0000211
12	.0000152	.0000431	.0000207
13	.0000131	.0000482	.0000206
14	.0000114	.0000532	.0000206
15	.0025602	.0027973	.0027077
16	.0037640	.0044205	.0041767
17	.0007366	.0007481	.0007855
18	.0007044	.0008982	.0005509
19	.0000219	.0000328	.0000218
20	.0000181	.0000379	.0000211
21	.0000152	.0000431	.0000207
22	.0000131	.0000482	.0000206
23	.0000114	.0000532	.0000206
24	.0025601	.0027974	.0027081
25	.0037639	.0044206	.0041768
26	.0007365	.0007481	.0007856
27	.0007040	.0008980	.0005506
28	.0000195	.0000345	.0000240
29	.0000164	.0000397	.0000231
30	.0000141	.0000449	.0000225
31	.0000124	.0000501	.0000222
32	.0000111	.0000551	.0000221
33	.0025904	.0027803	.0026603
34	.0038527	.0043212	.0041483
35	.0007508	.0007386	.0007723
36	.0020470	.0001197	.0015258
37	.0003265	.0008931	.0008897
38	.0003240	.0008958	.0008655
39	.0003225	.0009015	.0008484
40	.0003217	.0009084	.0008360
41	.0003212	.0009155	.0008266
42	.0022149	.0024915	.0022416
43	.0044939	.0052128	.0037693
44	.0011308	.0017629	.0013308
45	.0018591	.0036499	.0014959

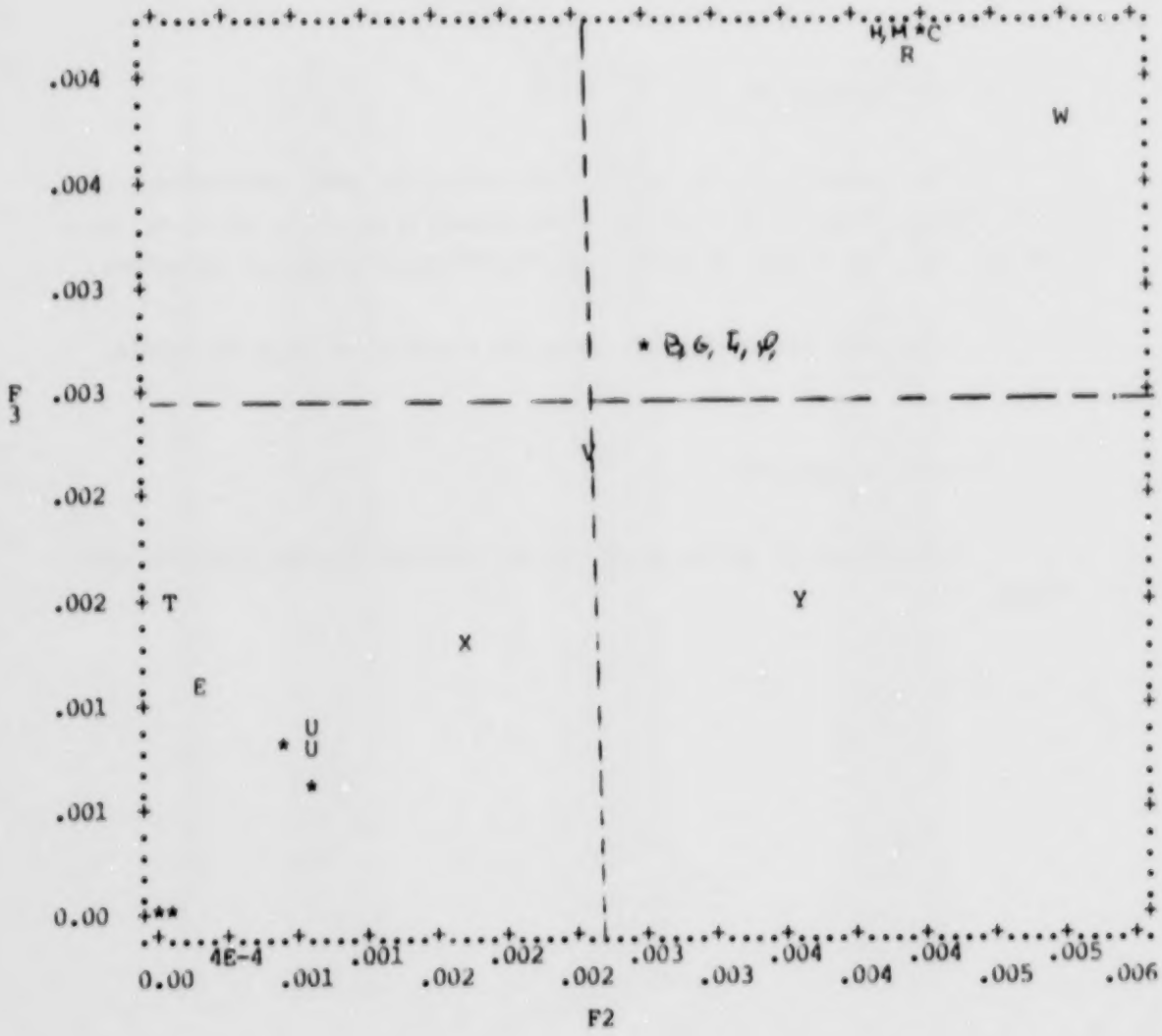
Graphe 11 : Représentation cartésienne des 45 solutions : type V (facteurs 1 et 2)



Graphe 11 (suite) : Représentation cartésienne des 45 solutions : type V (facteurs 1 et 3)



Graphe 11 (suite) : Représentation cartésienne des 45 solutions : type V (facteurs 2 et 3)



Des dix solutions aberrantes, quatre proviennent de la méthode de rotation promax, cinq de la méthode orthoblique et une de la méthode maxplane (il s'agit encore de la combinaison malifa (SMC) - maxplane).

Aucune distance n'est supérieure à 0,0053 pour ce type.

II.3.7 Le type VI

Tout comme pour le type V, dix solutions sont aberrantes : les cinq promax, B, G, L, Q, V et les cinq orthoblique, C, H, M, R, W. (voir tableau 16). Le graphe 12 permet une représentation de ces solutions.

Cependant toutes les distances se tiennent en deçà de 0,0048.

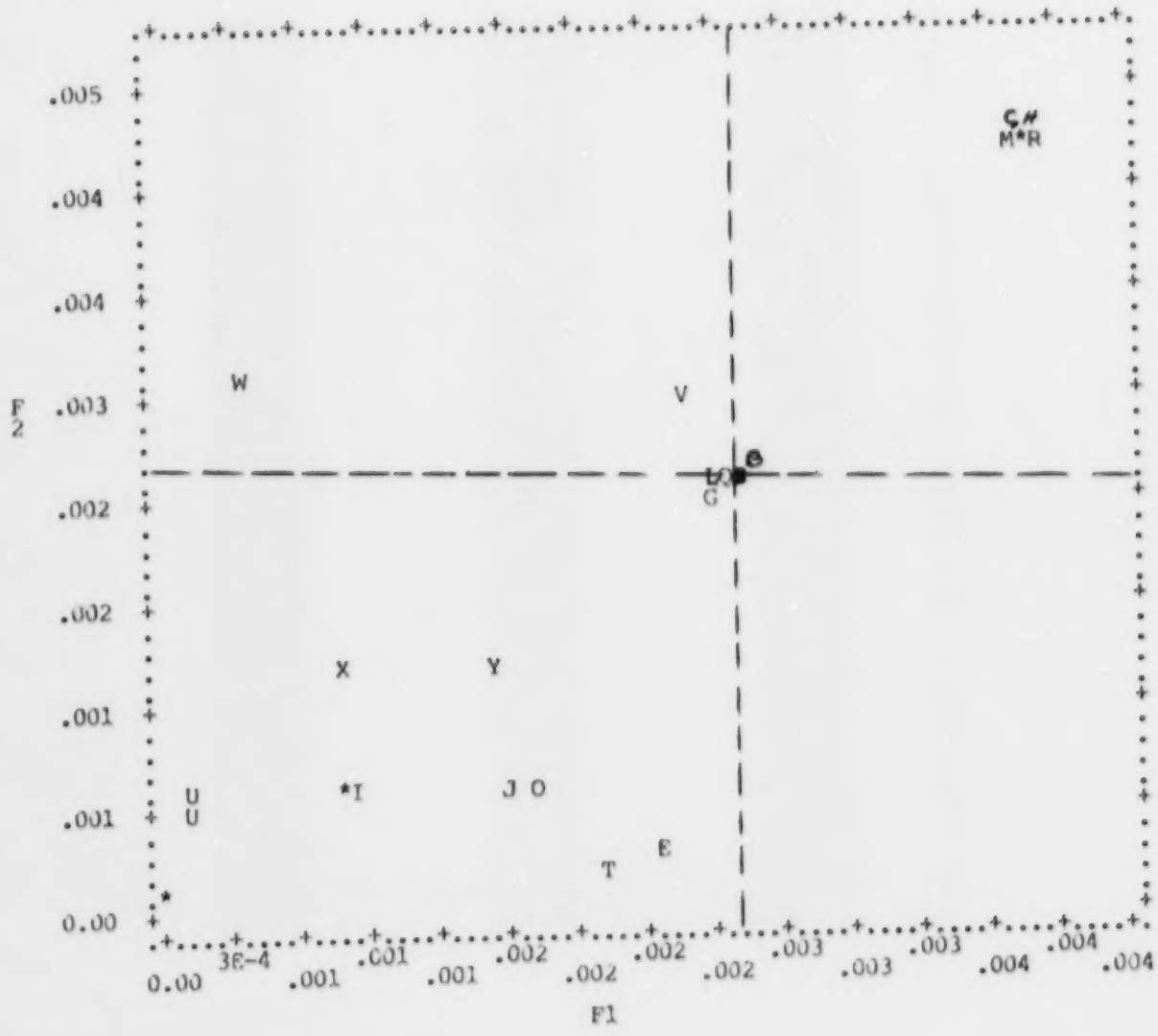
II.3.8 Le type VII

Le tableau 17 et le graphe 13 ne révèlent aucune solution aberrante.

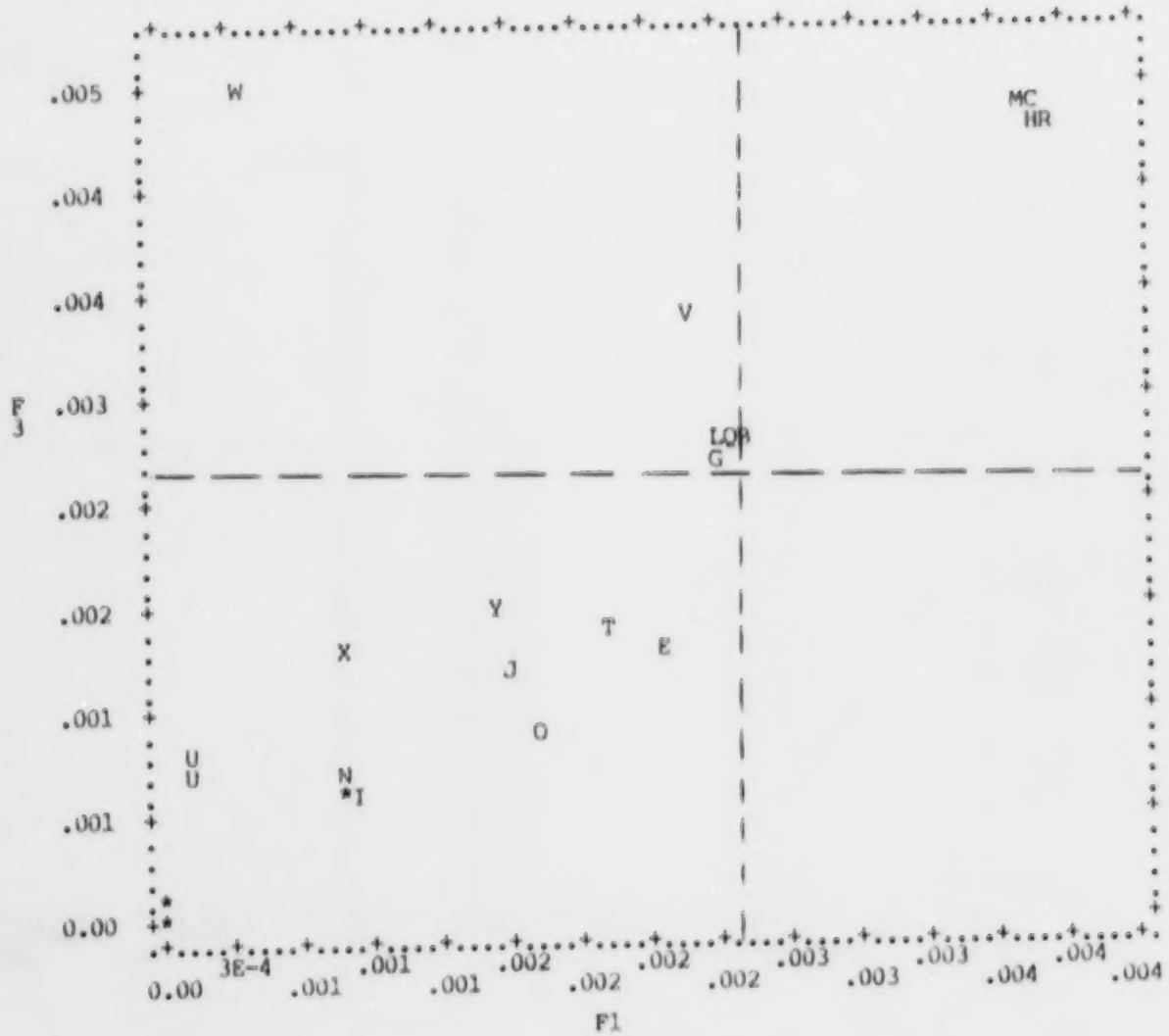
TABLEAU 16 : Distance des 45 solutions par rapport à la solution médiane : type VI

	FACTEUR 1	FACTEUR 2	FACTEUR 3
1	.0000046	.0000639	.0000892
2	.0000053	.0000736	.0000588
3	.0000084	.0000850	.0000388
4	.0000132	.0000971	.0000263
5	.0000192	.0001095	.0000192
6	.0025317	.0025236	.0027378
7	.0037725	.0044857	.0046310
8	.0008024	.0007639	.0007743
9	.0021341	.0003799	.0015585
10	.0000078	.0000847	.0001073
11	.0000081	.0000957	.0000754
12	.0000109	.0001081	.0000544
13	.0000156	.0001211	.0000412
14	.0000215	.0001343	.0000336
15	.0024136	.0023745	.0026292
16	.0037563	.0044397	.0046069
17	.0008128	.0007543	.0007750
18	.0015244	.0007260	.0014332
19	.0000073	.0000687	.0000908
20	.0000091	.0000789	.0000611
21	.0000131	.0000907	.0000418
22	.0000186	.0001033	.0000300
23	.0000251	.0001161	.0000235
24	.0024067	.0024972	.0027634
25	.0037123	.0044805	.0046479
26	.0007719	.0007747	.0007936
27	.0016440	.0006611	.0011289
28	.0000060	.0000654	.0000914
29	.0000065	.0000750	.0000606
30	.0000096	.0000862	.0000403
31	.0000144	.0000981	.0000276
32	.0000205	.0001102	.0000203
33	.0024792	.0024627	.0027081
34	.0038566	.0043919	.0045930
35	.0008027	.0007430	.0007718
36	.0019024	.0001965	.0016589
37	.0000905	.0006511	.0009332
38	.0000970	.0006695	.0008996
39	.0001036	.0006897	.0008778
40	.0001101	.0007101	.0008639
41	.0001164	.0007298	.0008553
42	.0022842	.0030422	.0034806
43	.0003374	.0031275	.0047908
44	.0008067	.0014242	.0016189
45	.0014299	.0013921	.0018302

Graphe 12 : Représentation cartésienne des 45 solutions : type VI (facteurs 1 et 2)



Graphe 12 (suite) : Représentation cartésienne des 45 solutions : type VI (facteurs 1 et 3)



Graphe 12 (suite) : Représentation cartésienne des 45 solutions : type VI (facteurs 2 et 3)

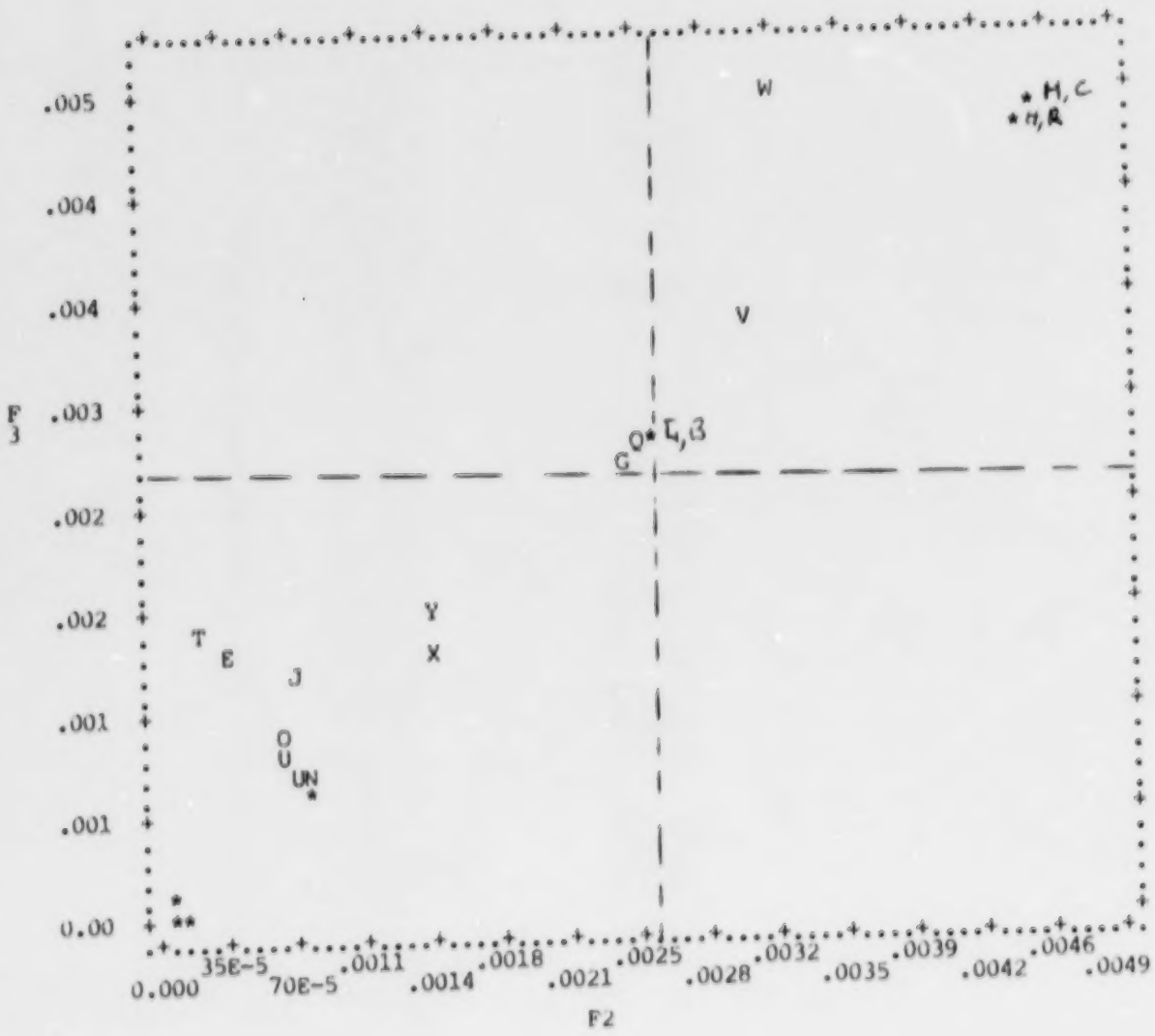
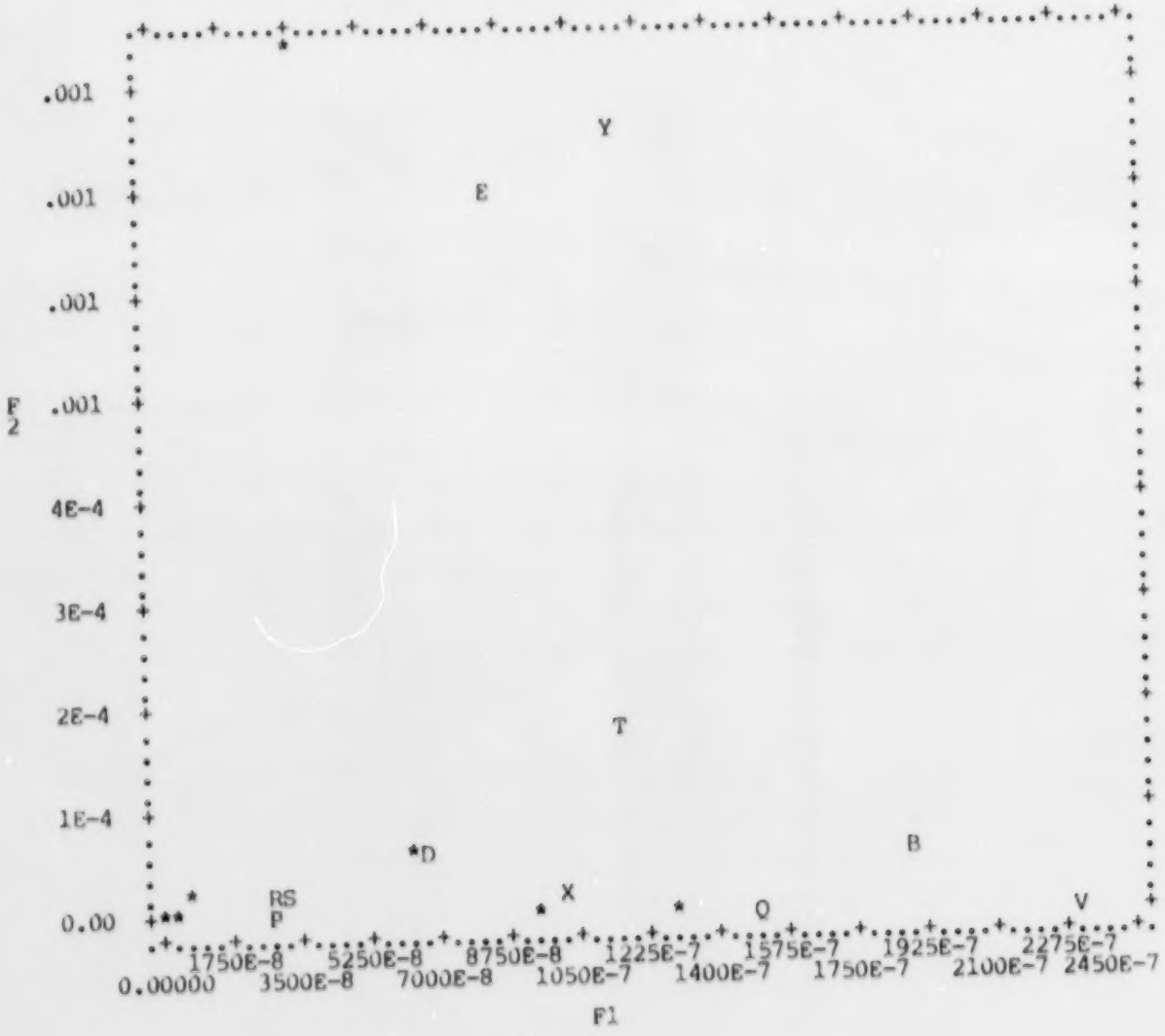


TABLEAU 17 : Distance des 45 solutions par rapport à la solution médiane : type VII

	FACTEUR 1	FACTEUR 2
1	.0000624	.0000534
2	.0000624	.0000534
3	.0000624	.0000534
4	.0000624	.0000534
5	.0000625	.0000534
6	.0001880	.0000578
7	.0000640	.0000540
8	.0000672	.0000610
9	.0000826	.0007010
10	.0000006	.0000090
11	.0000006	.0000090
12	.0000006	.0000090
13	.0000006	.0000090
14	.0000006	.0000091
15	.0001283	.0000086
16	.0000024	.0000096
17	.0000063	.0000174
18	.0000365	.0008305
19	.0000006	.0000090
20	.0000006	.0000090
21	.0000006	.0000090
22	.0000006	.0000090
23	.0000006	.0000091
24	.0001283	.0000086
25	.0000024	.0000096
26	.0000063	.0000174
27	.0000364	.0008304
28	.0000280	.0000089
29	.0000281	.0000089
30	.0000281	.0000089
31	.0000281	.0000090
32	.0000281	.0000090
33	.0001491	.0000095
34	.0000285	.0000118
35	.0000320	.0000167
36	.0001143	.0001799
37	.0000936	.0000044
38	.0000936	.0000044
39	.0000936	.0000044
40	.0000936	.0000044
41	.0000936	.0000044
42	.0002320	.0000058
43	.0000956	.0000062
44	.0001012	.0000128
45	.0001165	.0007519

Graphe 13 : Représentation cartésienne des 45 solutions : type VII



II.3.9 Le type VIII

Comme l'attestent le tableau 18 et le graphe 14, on ne retrouve aucune solution aberrante.

II.3.10 Sommaire des résultats

La première constatation majeure qui découle de ce critère de distance est que les matrices de corrélations des types III, IV, VII et VIII ne donnent pas de solutions aberrantes. Or ces matrices ont une particularité en commun : la faible intensité des corrélations.

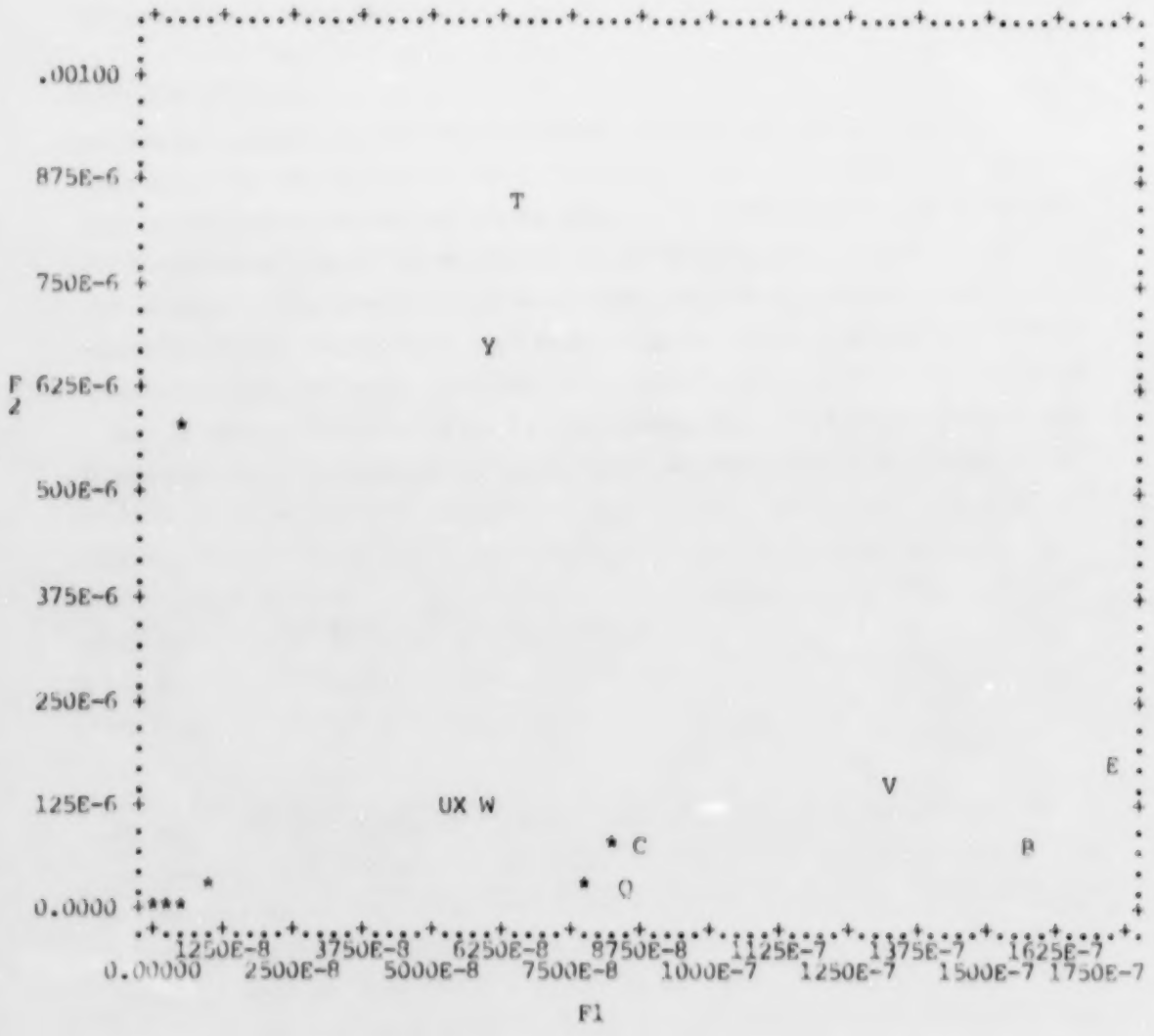
Soulignons que cette particularité comprend deux composantes. La première implique que les variables regroupées autour d'un même facteur sont plus faiblement correlées pour les types III, IV, VII, VIII que pour les autres types. La seconde implique que la corrélation moyenne entre, d'une part, les variables regroupées autour d'un premier facteur et, d'autre part, les variables regroupées autour d'un second facteur, est plus faible pour les types III, IV, VII, VIII que pour les autres types.

Seconde constatation : pour les matrices des types I, II, V et VI, qui donnent des solutions aberrantes, les méthodes de rotation semblent avoir une plus grande influence que les méthodes d'extraction en ce qui a trait à l'ampleur des distances. Par exemple la méthode de rotation orthoblique donne invariablement des solutions aberrantes pour ces quatre types de matrices et ce, quelle que soit la méthode d'extraction choisie. De même la méthode de rotation promax génère presque toujours des solutions aberrantes pour les matrices des types V et VI, même si les distances sont souvent près du seuil 0,0025. Une exception

TABLEAU 18 : Distance des 45 solutions par rapport à la solution médiane : type VIII

	FACTEUR 1	FACTEUR 2
1	.0000827	.0000697
2	.0000828	.0000697
3	.0000828	.0000697
4	.0000829	.0000697
5	.0000830	.0000697
6	.0001568	.0000793
7	.0000867	.0000725
8	.0000833	.0000718
9	.0001716	.0001635
10	.0000008	.0000073
11	.0000008	.0000073
12	.0000008	.0000073
13	.0000009	.0000073
14	.0000009	.0000074
15	.0000776	.0000138
16	.0000048	.0000100
17	.0000018	.0000101
18	.0000052	.0005631
19	.0000008	.0000073
20	.0000008	.0000073
21	.0000008	.0000073
22	.0000009	.0000073
23	.0000009	.0000074
24	.0000776	.0000138
25	.0000048	.0000100
26	.0000018	.0000101
27	.0000052	.0005633
28	.0000097	.0000207
29	.0000098	.0000208
30	.0000098	.0000208
31	.0000099	.0000208
32	.0000099	.0000209
33	.0000841	.0000322
34	.0000104	.0000244
35	.0000101	.0000230
36	.0000650	.0008505
37	.0000532	.0001252
38	.0000532	.0001252
39	.0000532	.0001252
40	.0000532	.0001253
41	.0000532	.0001253
42	.0001328	.0001394
43	.0000594	.0001305
44	.0000544	.0001276
45	.0000599	.0006817

Graphe 14 : Représentation cartésienne des 45 solutions : type VIII



toutefois : la méthode de rotation biquartimin qui donne des résultats très satisfaisants. Le type II est le seul cas où une méthode d'extraction paraît avoir joué un rôle prépondérant dans l'obtention de solutions aberrantes. En effet la méthode d'extraction malifa a donné des solutions aberrantes dans ce cas et ce, quelle que soit la méthode de rotation utilisée.

Une troisième et dernière constatation est de rigueur avant de terminer ce sommaire. Seuls les types I et II produisent des solutions aberrantes particulièrement éloignées de la solution de compromis (ce sont des solutions pour lesquelles un des d_i au moins est supérieur à 0,01) : pour le type I ce sont des solutions générées par la méthode de rotation maxplane ; pour le type II elles proviennent de la méthode d'extraction malifa. Ajoutons enfin que pour détecter des solutions spécialement aberrantes nous avons utilisé en plus des distances d_i un indice appelé le coefficient de congruence (Cureton et D'Agostino [1983, p. 392])

$$(c\,c)_i = \frac{\sum_{j=1}^{n} a_{ji}\,\bar{a}_{ji'}}{\sqrt{\sum_{j=1}^{n} a_{ji}^2 \sum_{j=1}^{n} \bar{a}_{ji}^2}}$$

pour le $i^{ième}$ facteur

Or seulement trois solutions, parmi l'ensemble des solutions des huit types, n'ont pu atteindre le seuil de congruence généralement accepté de 0,90. Il s'agit de trois solutions utilisant la méthode de rotation maxplane : les deux solutions A.F.P (MAXROW) - maxplane où $(cc)_1$ = 0,864 et A.F.P. (SMC) - maxplane où $(cc)_1$ = 0,887 pour le type I, ainsi que la solution malifa-maxplane où $(cc)_1$ = 0,898 pour le type II.

II.4 Critère 2 : le temps d'exécution informatique (C.P.U.)

II.4.1 Temps relatif aux méthodes d'extraction

Ce critère est fort différent des autres dans le sens qu'il ne fait pas référence à un aspect scientifique mais bien à un aspect bien pragmatique d'une analyse factorielle relié au coût d'une solution.

Pour chacune des 45 solutions, nous avons pris note, à partir des feuilles de sortie informatique, du temps d'exécution (appelé par l'ordinateur "CP seconds execution time") relatif aux différentes méthodes d'extraction employées. Le résultat se trouve au tableau 19.

Si on se concentre sur les deux premières lignes de ce tableau, on remarque que le temps d'exécution d'une A.C.P. est plus grand si on utilise le progiciel B.M.D.P. plutôt que S.P.S.S. et ce quel que soit le type de matrice de corrélations analysée. Le plus petit écart a été observé pour le type IV (10 variables, corrélations faibles, $\delta \neq 0$) ; il est d'un peu plus de deux secondes. C'est pour le type V (36 variables, 3 facteurs, corrélations fortes, $\delta = 0$) que nous avons constaté le plus grand écart : au-delà de 5 secondes !

Ces premières observations nous incitent à la prudence dans l'analyse des résultats ultérieurs relatifs à ce critère. Dans la suite nous ne comparerons donc que des méthodes provenant du même progiciel.

Les troisième et quatrième lignes du tableau peuvent aussi être comparées. La seule distinction entre les méthodes en question réside dans les estimés de communautés employés. Dans un cas on a utilisé MAXROW et dans l'autre SMC. Sauf pour le type V, où la différence

TABLEAU 19 : Temps d'exécution (en secondes) des méthodes d'extraction

TYPE / MÉTHODE	I	II	III	IV	V	VI	VII	VIII
A.C.P [1]	0,556	0,488	0,341	0,299	4,591	4,001	5,821	4,661
A.C.P. [2]	3,626	3,079	2,873	2,312	9,852	8,105	8,740	7,783
A.F.P (MAXROX) [2]	5,250	4,706	3,926	3,952	13,138	14,202	15,769	14,276
A.F.P (SMC) [2]	4,125	4,003	3,084	3,286	13,256	12,962	12,970	12,442
ALPHA (SMC) [1]	1,814	1,219	2,894	1,132	23,967	21,535	49,222	39,846
MALIFA (SMC) [2]	6,126	4,504	4,523	3,751	308,783	246,187	87,727	75,477

1 obtenue en utilisant le progiciel S.P.S.S.

2 obtenue en utilisant le progiciel B.M.D.P.

entre les temps d'exécution des deux méthodes est infime (≈ 0,12 secondes), la méthode A.F.P (MAXROW) prend plus de temps à exécuter les extractions de facteurs que la méthode A.F.P. (SMC). Cependant les écarts observés sont relativement petits : le plus faible, dans le cas du type IV, est de l'ordre de 0,67 secondes et le plus élevé, dans le cas du type VII (42 variables, corrélations faibles, $\delta = 0$), dépasse à peine 2,75 secondes.

D'un autre côté en comparant la méthode A.C.P. (version BMDP) à l'une ou l'autre des méthodes A.F.P, on constate que l'ampleur des écarts s'accroît sensiblement à mesure que le nombre de variables augmente de façon significative. En général, il peut être observé que les écarts sont plus faibles (de 0,21 à 1,64 seconde) pour les types I, II, III et IV dont le nombre de variables est petit (moins de 12). En effet dans le cas des types V, VI, VII et VIII, dont les matrices comportent beaucoup plus de variables (de 33 à 42), les écarts entre les temps d'exécution de la méthode A.C.P. et une des méthodes A.F.P. varient entre 3,29 et 7,03 secondes. Dans chacun de ces cas, c'est la méthode A.C.P. qui prend moins de temps pour exécuter une extraction de facteurs.

Par ailleurs, si on met en relief les temps d'exécution des méthodes A.C.P (version SPSS) et ALPHA (SMC), on peut faire des remarques analogues aux précédentes.

Alors que pour les types I, II, III et IV, les écarts sont faibles (moins de 2,6 secondes), ceux-ci sont beaucoup plus importants (de 17,5 à 43,4 secondes) pour les types V, VI, VII et VIII.

Une constatation du même ordre que les deux dernières mais encore plus claire ici, provient de la comparaison de la méthode malifa

(SMC) avec les autres méthodes obtenues à partir du progiciel BMDP. Les écarts entre les temps d'exécution de malifa et ces autres méthodes sont faibles (de 1,43 à 1,65 secondes) pour les types I, II, III et IV alors qu'ils sont réellement considérables (de 67,69 à 298,93 secondes) pour les types V, VI, VII et VIII. À chaque fois, la méthode A.C.P est la moins coûteuse en temps et malifa de loin la plus chronophage.

II.4.2 Temps relatif aux méthodes de rotation

Nous avons indiqué, au tableau 20, le temps d'exécution des diverses méthodes de rotation telles qu'utilisées après l'extraction des facteurs de l'une ou l'autre des méthodes d'extraction. C'est pourquoi ce tableau comporte 45 lignes correspondant aux 45 solutions obtenues. Insistons cependant sur le fait que le temps d'exécution inscrit dans ce tableau ne tient compte que des méthodes de rotation, excluant le temps d'exécution des méthodes d'extraction.

Il est à noter que pour les types I, II, III et IV, tous les temps d'excution se situent en deça d'une seconde. Aucun écart important ne peut donc être observé de ce côté. Une remarque similaire s'applique pour les types VII et VIII car la plus grande valeur n'est que de 1,149.

Les temps d'exécution les plus longs correspondent aux types V et VI. Ce qui n'est pas un résultat étonnant puisque ces types sont les seuls qui ont produit trois facteurs (au lieu de deux comme dans le cas de tous les autres types).

Or même ici les écarts entre les méthodes de rotation ne sont guère prononcés. L'écart le plus important est d'un peu plus de 2

TABLEAU 20 : Temps d'exécution (en secondes) des méthodes de rotation

SOLUTION \ TYPE	I	II	III	IV	V	VI	VII	VIII
1	0,513	0,455	0,513	0,420	1,054	1,036	0,599	0,532
2	0,498	0,458	0,478	0,418	1,101	1,036	0,614	0,557
3	0,484	0,458	0,450	0,420	1,109	1,096	0,615	0,662
4	0,515	0,446	0,465	0,400	1,063	1,054	0,572	0,534
5	0,512	0,469	0,452	0,437	1,098	1,079	0,585	0,566
6	0,730	0,736	0,637	0,702	1,852	1,597	1,011	0,892
7	0,663	0,674	0,697	0,684	1,949	1,544	1,002	0,888
8	0,695	0,684	0,632	0,729	1,859	1,672	1,029	0,863
9	0,965	0,823	0,900	0,964	2,645	2,731	1,073	1,059
10	0,480	0,476	0,443	0,414	1,107	1,023	0,556	0,569
11	0,472	0,458	0,467	0,422	1,071	1,045	0,616	0,565
12	0,461	0,437	0,516	0,437	1,059	1,054	0,592	0,581
13	0,519	0,477	0,518	0,422	1,056	1,057	0,583	0,630
14	0,524	0,477	0,496	0,455	1,063	1,040	0,613	0,626
15	0,667	0,658	0,675	0,740	1,863	1,643	1,081	0,870
16	0,682	0,702	0,702	0,715	1,979	1,554	0,949	0,930
17	0,728	0,628	0,645	0,719	1,764	1,719	1,081	0,889
18	0,873	0,810	0,822	0,885	2,882	2,700	1,074	1,050
19	0,508	0,436	0,472	0,427	1,070	1,061	0,573	0,607
20	0,471	0,507	0,445	0,424	1,071	1,062	0,595	0,615
21	0,482	0,418	0,468	0,413	1,095	1,033	0,602	0,556
22	0,511	0,429	0,493	0,411	1,120	1,025	0,573	0,551
23	0,541	0,481	0,464	0,427	1,064	1,090	0,561	0,559
24	0,698	0,690	0,675	0,694	1,869	1,597	1,047	0,893
25	0,700	0,669	0,651	0,684	1,811	1,513	0,962	0,845
26	0,720	0,647	0,630	0,708	1,780	1,726	0,951	0,987
27	0,953	0,786	0,953	0,904	2,687	2,493	1,078	1,046
28	0,492	0,453	0,484	0,419	1,084	0,999	0,596	0,621
29	0,457	0,439	0,422	0,426	1,056	1,001	0,551	0,603
30	0,471	0,434	0,449	0,422	1,042	0,999	0,553	0,566
31	0,461	0,426	0,437	0,437	1,031	1,007	0,558	0,551
32	0,479	0,449	0,472	0,438	1,042	1,043	0,560	0,545
33	0,719	0,689	0,692	0,691	1,923	1,573	1,040	0,884
34	0,721	0,628	0,669	0,713	1,806	1,553	0,945	0,890
35	0,685	0,654	0,648	0,669	1,754	1,724	0,949	0,986
36	0,980	0,778	0,899	0,859	2,454	2,569	1,149	0,976
37	0,456	0,432	0,530	0,423	1,043	0,999	0,573	0,557
38	0,447	0,443	0,452	0,420	1,012	1,012	0,586	0,565
39	0,470	0,420	0,532	0,397	1,023	1,022	0,585	0,566
40	0,473	0,444	0,472	0,393	1,076	1,009	0,571	0,597
41	0,475	0,460	0,494	0,426	1,076	1,088	0,572	0,567
42	0,732	0,738	0,681	0,676	1,009	1,568	0,991	0,849
43	0,694	0,685	0,698	0,646	1,822	1,489	0,917	0,928
44	0,668	0,619	0,645	0,673	1,837	1,794	1,036	0,951
45	0,984	0,884	0,916	0,855	3,079	2,177	1,129	1,048

secondes. Il correspond à la différence entre les temps d'exécution des méthodes de rotation malifa (le plus élevé) et orthomax ($\gamma = 0,25$) après extraction des facteurs selon la méthode maxplane.

Sauf en de très rares occasions c'est la méthode de rotation maxplane qui consomme le plus de temps d'exécution, et ce, quelle que soit la méthode d'extraction employée. D'ailleurs, on peut observer que les méthodes obliques sont plus chronophages que les méthodes orthogonales.

Nous avons tenu à relater cette dernière observation même si, tel que noté plus haut, les écarts sont plutôt minces entre les divers temps d'exécution.

II.4.3 Sommaire des résultats

Quatre constatations importantes ressortent de ces analyses relatives au temps d'exécution. Primo, le choix du progiciel peut faire varier le temps d'exécution d'une méthode d'extraction telle que l'analyse en composantes principales, suffisamment pour qu'un utilisateur éventuel prenne en compte ce choix. D'autant plus que les matrices de corrélations employées dans notre étude ne comprenaient qu'un nombre limité de variables et que pas plus de trois facteurs ont été extraits de ces matrices. Les données que nous avons en mains à cet effet, soit les deux premières lignes du tableau 19, nous permettent de soupçonner (à tout le moins!) que les écarts entre l'A.C.P. (version SPSS) et l'A.C.P. (version B.M.D.P) iraient croissants à mesure que le nombre de variables et de facteurs augmenteraient. D'autres recherches pourraient aussi étudier les variations de temps en utilisant d'autres progiciels (v.g S.A.S....) et aussi d'autres méthodes d'extraction (v.g A.F.P....)

Secundo, il appert que ce sont surtout les méthodes d'extraction qui font varier les temps d'exécution.

Quant aux méthodes de rotation les écarts entre les temps d'exécution semblent vraiment trop minces pour s'en préoccuper. Du moins si on ne considère que des matrices de corrélations telles que celles utilisées ici. Il serait cependant plus prudent, à notre avis, d'effectuer des études similaires utilisant des données plus complexes, c'est-à-dire avec plus de facteurs. Peut-être alors observera-t-on plus nettement la tendance déjà mentionnée qu'ont les méthodes obliques à consommer plus de temps ! La méthode maxplane devrait faire l'objet d'une attention toute particulière.

Tertio, on a vu que les écarts entre les temps d'exécution des méthodes d'extractions pour les types I, II, III ou IV étaient négligeables comparativement à ceux observés pour les types V, VI, VII ou VIII, autant pour les méthodes du progiciel SPSS que du progiciel BMDP.

Toutes les comparaisons effectuées entre les diverses méthodes d'extraction d'un même progiciel, sauf A.F.P (MAXROW) versus A.F.P. (SMC), nous mènent en effet à cette remarque. Ainsi, lorsque le nombre de variables est grand (plus de 30), il devient donc beaucoup plus important de choisir judicieusement la bonne méthode d'extraction, en tout cas si les questions économiques sont prises en compte.

Quarto, nous avons trouvé les méthodes alpha et (surtout) malifa particulièrement dispendieuses dans les cas où plusieurs variables étaient en jeu (types V, VI, VII et VIII). Il nous paraît même que le qualificatif "luxueuse" n'est pas exagéré pour y référer. C'est pourquoi nous croyons qu'un utilisateur sensible un tant soit peu au coût d'une recherche devrait avoir de sérieuses raisons de les employer.

II.5 Critère 3 : l'invariance factorielle

II.5.1 Description

Ce critère nous a été suggéré par Kaiser [1958]. Nous l'avons décrit la première fois dans Bertrand et Dufour [1981]. Il s'agit d'invariance factorielle psychométrique. C'est-à-dire que nous voulons mesurer jusqu'à quel point l'échantillonnage des variables (d'un même domaine d'étude) peut influencer la solution factorielle.

De façon opérationnelle, il s'agit de mesurer la distance entre des patterns factoriels provenant d'ensembles de variables distincts mais représentant le même domaine d'étude. Ceci est accompli en choisissant d'abord un certain nombre de variables pour chaque groupe (ou dimension ou facteur). Puis, après avoir calculé la matrice de corrélations de ces variables, on obtient le pattern factoriel correspondant à l'aide d'une méthode d'extraction et d'une méthode de rotation choisies au préalable. Ensuite on ajoute des variables, possiblement dans chaque groupe, on calcule la nouvelle matrice de corrélations et on obtient un nouveau pattern factoriel à l'aide de la même méthode d'extraction et de la même méthode de rotation. On itère cette procédure aussi souvent que requis et enfin on compare les différents patterns factoriels. Plus ceux-ci sont près les uns des autres, plus on pourra conclure que la solution possède une invariance factorielle.

Le tableau 21 donne une indication de ce que nous avons obtenue en utilisant la méthode d'extraction A.C.P et la méthode de rotation orthomax ($\gamma = 0$) pour le type I. Pour chacun des deux groupes (facteurs) nous pouvons voir l'évolution des saturations du pattern factoriel suivant qu'il y avait quatre, six, huit, dix ou douze variables

TABLEAU 21 : Évolution des saturations des deux facteurs pour la solution 1 du type I.

TYPE I

METHODE:PFA ROTATION:OBLIQUE GAMMA=0,00

FACTEUR 1

	4	6	8	10	12
V1	82	55	34	24	8
V2	962	927	903	876	866
V9	264	271	269	281	270
V11	955	964	965	965	961
V5		404	415	439	432
V24		914	922	929	926
V6			929	944	942
V20			413	440	433
V10				508	505
V12				882	882
V15					-487
V17					234

FACTEUR 2

	4	6	8	10	12
V1	973	941	914	887	861
V2	123	91	68	26	12
V9	940	954	953	942	936
V11	199	215	217	193	195
V5		878	891	891	899
V24		308	319	307	317
V6			314	303	314
V20			875	878	886
V10				811	829
V12				385	405
V15					224
V17					-535

$VF_1 = 0,575 \times 10^{-4}$

$VF_2 = 0,859 \times 10^{-4}$

dans la matrice de corrélations analysée. Ainsi le nombre 82 (lire : 0,082) correspond à la saturation du facteur 1 pour la variable V_1 obtenue après avoir analysé une matrice initiale des corrélations de quatre variables. Cette saturation s'est transformée successivement en 55, 34, 24 puis 8 suivant qu'on ajoutait deux, quatre, six ou huit variables aux quatre déjà présentes dans la matrice initiale.

Nous avons dû construire un indice afin d'évaluer la distance entre ces différents patterns factoriels. Il a été appelé l'indice de variation factorielle et défini :

$$VF_p = \frac{1}{4} \frac{\sum_{n \in X} \sum_{j=1}^{n+\delta} \left(a_{j(n+\delta)p} - a_{j(n+\delta+2)p} \right)^2}{d}$$

où p = nombre de facteurs

$X = \left\{ 2x + 4 \mid x \in \mathbb{N},\ 2x + 6 + \delta \leq v \right\}$

v = nombre total de variables considérées

$\delta = n_1 - n_2$, où n_1 (resp. n_2) est le nombre de variables du premier (resp. second) groupe,

$a_{j(n+\delta)p}$ = saturation du $p^{\text{ième}}$ facteur et de la $j^{\text{ème}}$ variable, obtenu de l'analyse d'une matrice de corrélations de dimension n , pour un δ spécifié au préalable

d = nombre de termes (ou de différences) du numérateur.

Cet indice étant calculé à partir d'un tableau comme le tableau 21, la $j^{\text{ième}}$ variable correspond à la $j^{\text{ième}}$ ligne de ce tableau.

L'indice VF_p varie de 0 à 1 car si toutes les saturations comparées sont identiques,
c.à.d. $a_{j(n+\delta)p} = a_{j(n+\delta+2)p}$ alors $VF_p = 0$.

De plus la différence maximale entre deux saturations étant 2 (dans le cas où une des saturations comparées vaut 1 et l'autre -1), la valeur maximale de

$$VF_p \text{ devient } \frac{\frac{1}{4}\ d.2^2}{d} = 1.$$

En conséquence plus cet indice est petit, plus on note un haut degré d'invariance factorielle. Inversement plus cet indice est grand, plus on note un faible degré d'invariance factorielle. De façon plus précise, nous dirons qu'une solution manque d'invariance factorielle si $VF_p \geqslant 25 \times 10^{-4}$ pour au moins un "p".

Comme $VF_p = \frac{\sum\sum (0,1)^2}{4d} = \frac{(0,01)\ N}{4N} = 0,0025 = 25 \times 10^{-4}$

le seuil 25×10^{-4} correspond à une différence moyenne entre les saturations de 0,1

Nous noterons tout particulièrement les solutions où $VF_p \geqslant 100 \times 10^{-4}$, c'est-à-dire celles dont la différence moyenne entre les saturations est d'au moins 0,2.

Ces seuils correspondent à ceux établis plus haut en II.3.

À l'aide du tableau 21, il est possible de montrer comment se calculent ces indices. Pour le premier facteur on voit que $VF_1 = 0,575 \times 10^{-4}$. Or ce nombre correspond au résultat des calculs suivants. On effectue d'abord les sommes :

$$\sum\sum = (0,082 - 0,055)^2+(0,962 - 0,927)^2+(0,264 - 0,271)^2+(0,955 - 0,964)^2$$

$$+ (0,055 - 0,034)^2+(0,927 - 0,903)^2+ \text{----------} +(0,914 - 0,922)^2$$

$$+$$

$$"$$

$$"$$

$$"$$

$$+ (0,024 - 0,008)^2+ \text{----------} +(0,882 - 0,882)^2$$

Puis on calcule le nombre de termes de cette somme : ici d = 28. Enfin on divise la somme $\sum\sum$ par 4d. Le résultat donne $0,575 \times 10^{-4}$.

Le même procédé est utilisé pour le calcul de VF_2.

II.5.2 Le type I

Le tableau 22 présente les indices VF_1 et VF_2 pour les 45 solutions du type I. Tel qu'indiqué tous les nombres doivent être multipliés par 10^{-4}.

On voit que seulement quatre solutions manquent d'invariance factorielle car l'indice VF_1 dépasse le seuil 25 ($\times 10^{-4}$) pour les solutions 9, 18, 27 et 43. Trois de ces solutions proviennent de la méthode de rotation maxplane, les méthodes d'extraction étant respectivement A.C.P., A.F.P (MAXROW) et A.F.P (SMC). L'autre solution en cause est issue de la combinaison malifa-orthoblique.

TABLEAU 22 : Indice de variation factorielle ($\times 10^{-4}$) pour le type I

	Solution	Facteur 1 (VF_1)	Facteur 2 (VF_2)
A.C.P.	1	0,575	0,859
	2	0,492	0,716
	3	0,492	0,685
	4	0,508	0,680
	5	0,524	0,683
	6	7,141	7,004
	7	5,354	3,813
	8	3,226	3,133
	9	68,024	23,444
A.F.P. (MAXROW)	10	0,841	1,212
	11	0,775	1,098
	12	0,777	1,076
	13	0,791	1,073
	14	0,806	1,075
	15	7,331	6,882
	16	5,657	4,440
	17	3,431	3,373
	18	39,438	13,567
A.F.P. (SMC)	19	0,968	1,164
	20	0,870	1,050
	21	0,861	1,029
	22	0,870	1,027
	23	0,883	1,030
	24	7,325	6,569
	25	5,713	4,281
	26	3,326	3,240
	27	25,178	12,920
ALPHA (SMC)	28	1,436	2,025
	29	1,360	1,951
	30	1,353	1,943
	31	1,362	1,947
	32	1,373	1,951
	33	7,950	7,060
	34	6,242	5,181
	35	3,823	3,929
	36	12,302	9,082
MALIFA (SMC)	37	3,669	4,709
	38	3,070	3,305
	39	2,805	3,538
	40	2,667	3,343
	41	2,587	3,228
	42	6,353	12,354
	43	32,564	5,275
	44	3,079	8,389
	45	6,042	13,845

II.5.3 Le type II

D'après le tableau 23, un grand nombre de solutions manquent d'invariance factorielle pour le type II. Ceci corrabore d'ailleurs des résultats déjà obtenus par Bertrand et Dufour [1981].

On dénombre en effet 24 solutions fautives. Pour les méthodes d'extraction A.C.P., A.F.P (MAXROW) et A.F.P. (SMC) on retrouve cinq solutions dont VF_1 ou VF_2 dépasse le seuil de 25 ($\times 10^{-4}$). Dans chacun de ces trois cas les cinq mêmes méthodes de rotation sont en cause, il s'agit de orthomax ($\gamma = 0,75$) orthomax ($\gamma = 1$) ou varimax, promax, orthoblique et biquartimin. Dans le cas de la méthode d'extraction alpha, six solutions manquent d'invariance ; les méthodes de rotation correspondantes sont orthomax ($\gamma = 0,75$), orthomax ($\gamma = 1$) ou varimax, promax, orthoblique, biquartimin, maxplane.

Les trois dernières solutions proviennent de la méthode d'extraction malifa : les méthodes de rotation en cause sont promax, orthoblique et maxplane. Dans ce dernier cas, il est à souligner que l'indice dépasse même 100×10^{-4}.

Remarquons en terminant qu'aucune solution issue des méthodes de rotation promax et orthoblique, n'a pu être considérée comme invariante.

TABLEAU 23 : Indice de variation factorielle ($\times 10^{-4}$) pour le type II

Solution	Facteur 1 (VF_1)	Facteur 2 (VF_2)
1	4,721	10,956
2	7,123	15,596
3	10,259	21,460
4	13,725	27,562
5	17,067	32,987
6	33,846	8,457
7	39,989	11,647
8	34,654	9,983
9	13,995	8,886
10	3,492	9,114
11	5,489	13,453
12	8,323	19,336
13	11,591	25,664
14	14,807	31,370
15	31,078	6,259
16	38,772	9,335
17	31,191	7,659
18	12,102	5,913
19	3,995	11,090
20	6,391	15,868
21	9,473	21,788
22	12,830	27,836
23	16,019	33,106
24	33,016	8,375
25	38,953	11,914
26	33,571	9,985
27	11,887	6,500
28	8,281	12,700
29	11,176	17,434
30	14,757	23,242
31	18,570	29,142
32	22,146	34,262
33	34,701	6,972
34	43,150	9,990
35	39,491	10,672
36	66,298	16,376
37	4,105	14,875
38	2,521	11,446
39	1,582	9,534
40	1,324	8,987
41	1,588	9,320
42	12,456	26,710
43	16,915	26,474
44	4,567	14,861
45	111,842	34,028

II.5.4 Le type III

Comme on peut le constater au tableau 24 toutes les solutions du type III doivent être considérées comme ayant un degré suffisant d'invariance factorielle.

II.5.5 Le type IV

Seulement une solution manque d'invariance factorielle ici. Le tableau 25 montre que cette solution provient de la combinaison malifa-maxplane.

À noter aussi que le facteur 1 paraît beaucoup plus invariant que le facteur 2.

II.5.6 Le type V

Neuf solutions comportent au moins un indice VF_p plus grand que le seuil 25 ($\times 10^{-4}$). D'après le tableau 26 on peut voir qu'il s'agit de toutes les solutions issues de la méthode d'extraction malifa. La combinaison malifa-maxplane comporte des indices VF_2 et VF_3 particulièrement élevés.

II.5.7 Le type VI

Tout comme pour le type V, les neuf solutions de la méthode d'extraction malifa manquent d'invariance factorielle (tableau 27). De plus trois autres solutions sont dans cette catégorie ; elles ont été produites par la méthode de rotation maxplane.

TABLEAU 24 : Indice de variation factorielle ($\times 10^{-4}$) pour le type III

Solution	Facteur 1 (VF_1)	Facteur 2 (VF_2)
1	1,371	2,781
2	1,381	2,802
3	1,391	2,823
4	1,400	2,844
5	1,410	2,864
6	1,519	4,328
7	1,291	4,564
8	1,281	3,452
9	7,379	6,090
10	2,306	6,155
11	2,320	6,172
12	2,333	6,189
13	2,346	6,206
14	2,358	6,223
15	2,894	7,746
16	2,474	7,990
17	2,255	6,822
18	11,518	8,040
19	2,297	6,136
20	2,310	6,153
21	2,324	6,170
22	2,337	6,187
23	2,349	6,204
24	2,884	7,727
25	2,464	7,970
26	2,246	6,802
27	10,035	14,367
28	2,545	7,206
29	2,557	7,230
30	2,570	7,253
31	2,583	7,276
32	2,595	7,299
33	3,240	9,220
34	2,837	9,312
35	2,542	7,957
36	9,824	10,101
37	1,696	3,691
38	1,711	3,720
39	1,727	3,750
40	1,744	3,779
41	1,762	3,809
42	2,476	5,341
43	2,023	4,536
44	1,620	4,295
45	2,701	3,820

TABLEAU 25 : Indice de variation factorielle ($\times 10^{-4}$) pour le type IV

Solution	Facteur 1 (VF_1)	Facteur 2 (VF_2)
1	1,036	2,791
2	1,047	2,799
3	1,063	2,819
4	1,081	2,850
5	1,103	2,893
6	1,353	4,815
7	1,084	5,689
8	1,062	3,431
9	0,949	2,948
10	1,847	6,762
11	1,850	6,759
12	1,858	6,766
13	1,872	6,782
14	1,890	6,805
15	2,889	10,204
16	2,202	9,470
17	1,950	7,381
18	3,793	19,488
19	1,847	6,868
20	1,851	6,866
21	1,859	6,872
22	1,873	6,888
23	1,891	6,911
24	2,895	10,318
25	2,206	9,580
26	1,952	7,489
27	4,306	14,754
28	1,593	11,904
29	1,594	11,897
30	1,598	11,898
31	1,606	11,908
32	1,618	11,927
33	2,930	16,927
34	2,242	15,903
35	1,817	12,666
36	3,460	13,575
37	1,709	14,358
38	1,723	14,353
39	1,744	14,357
40	1,772	14,370
41	1,807	14,389
42	3,012	17,271
43	2,554	15,272
44	1,902	15,059
45	4,167	25,121

TABLEAU 26 : Indice de variation factorielle ($\times 10^{-4}$) pour le type V

Solution	Facteur 1 (VF_1)	Facteur 2 (VF_2)	Facteur 3 (VF_3)
1	1,703	2,223	1,458
2	1,735	2,263	1,465
3	1,766	2,300	1,471
4	1,796	2,335	1,477
5	1,824	2,369	1,483
6	3,217	5,083	4,499
7	4,240	5,954	4,433
8	2,155	3,191	1,895
9	7,201	5,729	10,809
10	1,948	2,429	2,047
11	1,979	2,466	2,053
12	2,009	2,502	2,059
13	2,038	2,535	2,064
14	2,065	2,566	2,069
15	3,506	5,456	5,320
16	4,533	6,261	5,065
17	2,396	3,389	2,477
18	8,734	8,879	12,174
19	1,925	2,266	2,004
20	1,957	2,304	2,010
21	1,987	2,339	2,016
22	2,015	2,373	2,021
23	2,043	2,404	2,026
24	3,447	5,290	5,269
25	4,484	6,073	4,999
26	2,366	3,249	2,435
27	7,417	6,217	9,638
28	3,432	4,308	4,126
29	3,462	4,343	4,133
30	3,491	4,376	4,141
31	3,519	4,407	4,148
32	3,546	4,437	4,154
33	5,085	7,351	7,102
34	5,958	8,099	7,025
35	3,841	5,087	4,467
36	9,975	8,391	17,666
37	11,641	31,120	16,573
38	11,573	30,846	16,500
39	11,533	30,647	16,444
40	11,510	30,496	16,399
41	11,496	30,377	16,364
42	14,081	49,540	22,990
43	26,279	49,359	38,449
44	12,157	34,771	17,051
45	18,915	85,829	83,840

TABLEAU 27 : Indice de variation factorielle ($\times 10^{-4}$) pour le type VI

Solution	Facteur 1 (VF_1)	Facteur 2 (VF_2)	Facteur 3 (VF_3)
1	2,277	2,112	0,593
2	2,303	2,142	0,623
3	2,336	2,182	0,653
4	2,374	2,228	0,682
5	2,415	2,279	0,710
6	3,641	6,403	5,964
7	5,053	7,239	6,093
8	2,733	3,503	1,438
9	4,385	23,797	13,679
10	3,088	2,647	0,866
11	3,124	2,700	0,898
12	3,165	2,759	0,930
13	3,210	2,822	0,960
14	3,257	2,887	0,988
15	4,641	7,078	6,945
16	6,151	7,918	6,476
17	3,562	4,045	1,718
18	7,028	40,806	9,245
19	3,531	2,577	0,808
20	3,545	2,602	0,839
21	3,567	2,639	0,871
22	3,594	2,684	0,902
23	3,626	2,733	0,931
24	5,084	7,061	6,886
25	6,678	7,767	6,350
26	4,015	3,984	1,649
27	8,947	41,497	8,649
28	3,150	2,858	1,739
29	3,170	2,887	1,752
30	3,197	2,926	1,768
31	3,229	2,971	1,785
32	3,265	3,021	1,802
33	4,707	7,359	6,959
34	6,187	7,318	6,867
35	3,589	4,036	2,426
36	7,592	27,928	9,555
37	29,945	44,425	17,657
38	29,909	43,862	17,388
39	29,909	43,449	17,197
40	29,932	43,135	17,060
41	29,972	42,888	16,961
42	40,396	64,129	24,125
43	43,460	71,171	32,400
44	32,692	49,700	17,727
45	54,488	79,647	49,212

TABLEAU 28 : Indice de variation factorielle ($\times 10^{-4}$) pour le type VII

Solution	Facteur 1 (VF_1)	Facteur 2 (VF_2)
1	0,842	1,119
2	0,842	1,120
3	0,843	1,120
4	0,843	1,121
5	0,843	1,121
6	1,001	1,363
7	0,862	1,359
8	0,832	1,298
9	2,476	2,429
10	0,853	0,664
11	0,853	0,665
12	0,854	0,665
13	0,854	0,666
14	0,855	0,666
15	0,991	0,882
16	0,848	0,875
17	0,831	0,828
18	1,448	2,844
19	0,835	0,708
20	0,835	0,708
21	0,835	0,709
22	0,836	0,709
23	0,836	0,710
24	0,968	0,919
25	0,829	0,917
26	0,814	0,870
27	1,153	2,896
28	0,962	0,718
29	0,962	0,718
30	0,962	0,718
31	0,963	0,719
32	0,963	0,719
33	1,095	0,983
34	1,224	1,012
35	0,942	0,888
36	2,947	3,241
37	0,344	0,944
38	0,345	0,944
39	0,345	0,945
40	0,346	0,945
41	0,347	0,946
42	0,413	1,083
43	0,360	0,981
44	0,302	1,104
45	0,383	3,144

II.5.8 Le type VII

Aucune solution n'a un indice de variation factorielle supérieure à 25 ($\times 10^{-4}$) pour ce type (tableau 28).

II.5.9 Le tupe VIII

Tout comme pour le type VII, aucune solution de ce type n'a un indice de variation factorielle qui dépasse le seuil de 25 (tableau 29).

II.5.10 Sommaire des résultats

Comme première observation, on note qu'en général les solutions produites par les méthodes de rotation orthogonales tendent à être plus invariantes (VF_p plus petits) que les solutions obliques.

Seconde observation : les solutions des types dont les matrices ont des corrélations faibles, soit les types III, IV, VII et VIII semblent beaucoup plus invariantes. Un seul cas de solution ayant un indice plus grand que 25 a été constaté pour l'ensemble de ces 180 (4 x 45) solutions. Par contraste pour les types I, II, V et VI, 49 solutions manquent d'invariance factorielle.

Troisième observation : pour chaque type comportant des solutions où il y a un manque d'invariance factorielle, au moins une de ces solutions provient de la méthode d'extraction malifa et au moins une de la méthode de rotation maxplane. La combinaison malifa-maxplane donnent des solutions tout particulièrement déviantes pour le type II et dans

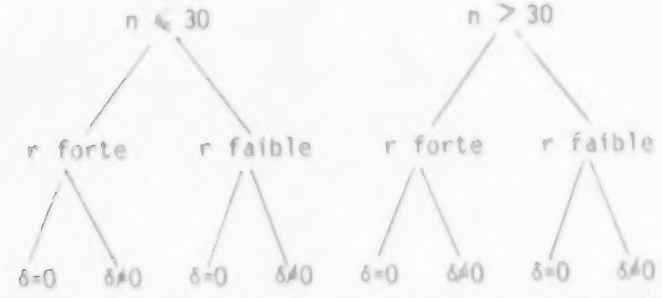
TABLEAU 29 : Indice de variation factorielle ($\times 10^{-4}$) pour le type VIII

Solution	Facteur 1 (VF_1)	Facteur 2 (VF_2)
1	1,412	1,680
2	1,422	1,694
3	1,434	1,709
4	1,448	1,726
5	1,465	1,744
6	1,827	2,122
7	1,504	2,296
8	1,391	2,059
9	4,210	4,309
10	1,062	1,500
11	1,075	1,511
12	1,090	1,525
13	1,108	1,541
14	1,129	1,559
15	1,520	1,931
16	1,099	2,099
17	1,020	1,865
18	1,947	2,500
19	1,059	1,500
20	1,071	1,511
21	1,087	1,525
22	1,105	1,541
23	1,126	1,559
24	1,516	1,932
25	1,095	2,100
26	1,017	1,865
27	1,946	2,624
28	1,263	1,447
29	1,274	1,460
30	1,288	1,475
31	1,304	1,492
32	1,323	1,511
33	1,700	1,907
34	1,482	2,425
35	1,233	1,824
36	3,441	1,726
37	1,833	3,936
38	1,849	3,948
39	1,867	3,962
40	1,888	3,978
41	1,911	3,996
42	2,266	4,289
43	2,038	4,310
44	1,786	4,292
45	2,425	4,397

une moindre mesure pour les types IV, V et VI. On notera aussi qu'aucune des solutions "malifa" des types V et VI ne peut être considérée comme invariante.

Quatrième et dernière observation : les indices de variation factorielle VF_p des types où $\delta = 0$, soit les types I, III, V, VII, sont en général plus faibles que ceux des types correspondants où $\delta \neq 0$, soit les types II, IV, VI, VIII. Ce résultat recoupe celui obtenu par Bertrand et Dufour [1981]. D'ailleurs il y a toujours autant, sinon plus, de solutions des types II, IV, VI, VIII qui manquent d'invariance factorielle que des types correspondants I, III, V, VII. Le tableau 30 résume la situation.

Tableau 30 : Nombre (Nb) de solutions qui manquent d'invariance factorielle



TYPE	I	II	III	IV	V	VI	VII	VIII
Nb	4	24	0	1	9	12	0	0

II.6 Critère 4 : Le jugement des spécialistes

Nous avons demandé à trois spécialistes reconnus de l'analyse factorielle d'indiquer leur(s) choix personnel(s) de meilleure(s) solution(s) pour chacun des types I, III, V et VII.

Il s'agit de MM. Djavid Ajar, professeur à l'Université de Montréal, Jean-Jacques Bernier, professeur à l'Université Laval et Horacio Rimoldi, directeur du Centre Interdisciplinario de Investigaciones en Psicologia Matematica y Experimental de Buenos Aires en Argentine.

Dans la suite du texte, nous ferons référence aux juges 1, 2 et 3 pour désigner ces spécialistes[1].

II.6.1 Le type I

Après avoir étudié le graphe du pattern factoriel de la solution A.F.P (MAXROW)-promax, le juge 1 conclut qu'il s'agit là "d'une mauvaise structure et que rien ne peut être fait" au niveau de rotations manuelles supplémentaires.

Le juge 2 souligne que "toutes les solutions lui paraissent assez semblables" mais que "les solutions suivantes (dans l'ordre) sont meilleures :"

---

1 L'ordre des juges ne correspond pas à l'ordre alphabétique des spécialistes donné plus haut.

malifa-biquartimin
A.C.P.-biquartimin
alpha-promax
A.F.P(SMC)-promax
A.F.P(MAXROW)-orthoblique
alpha-orthoblique
A.C.P-orthoblique

Le juge 3 opte pour une A.C.P. ou une A.F.P(MAXROW) comme méthode d'extraction et une orthomax comme méthode de rotation.

II.6 2 Le type III

Le juge 1 indique que la solution A.F.P(MAXROW)-promax est bonne ici et ne demande aucune autre rotation.

Le juge 2 choisit comme meilleures solutions les cinq A.C.P. orthomax parmi les solutions orthogonales et A.C.P(maxplane), A.F.P-(MAXROW)-biquartimin, A.F.P(MAXROW)-maxplane, A.F.P(SMC)-biquartimin et A.F.P(SMC)-maxplane, alpha-maxplane et malifa-maxplane parmi les solutions obliques. Ce juge affirme qu'encore ici "les solutions sont assez semblables".

Le juge 3 choisit dans l'ordre les solutions A.F.P(SMC)-biquartimin et A.F.P(SMC)-varimax.

II.6.3 Le type V

Le juge 1 mentionne que la solution privilégiée pour ce type consiste en une version modifiée de la solution A.F.P(MAXROW)-promax à

l'aide de rotations manuelles. Comme il s'agit d'une solution peu connue, nous la reproduisons (du moins le graphe du pattern factoriel) au graphe 15.

Le juge 2 précise "qu'il est très difficile que l'on identifie la meilleure solution" dans ce cas. En contrepartie il indique quelles solutions semblent "moins parcimonieuses". Il s'agit des cinq solutions orthogonales malifa-orthomax et des solutions obliques ; A.C.P-biquartimin, A.C.P-maxplane, A.F.P(MAXROW)-biquartimin, A.F.P(MAXROW)-maxplane, A.F.P(SMC)-biquartimin, A.F.P(SMC)-maxplane, alpha-biquartimin, alpha-maxplane.

Le juge 3 opte pour les solutions A.F.P(SMC)-biquartimin, A.F.P(SMC)-varimax, alpha-biquartimin et alpha-varimax.

II.6.4 Le type VII

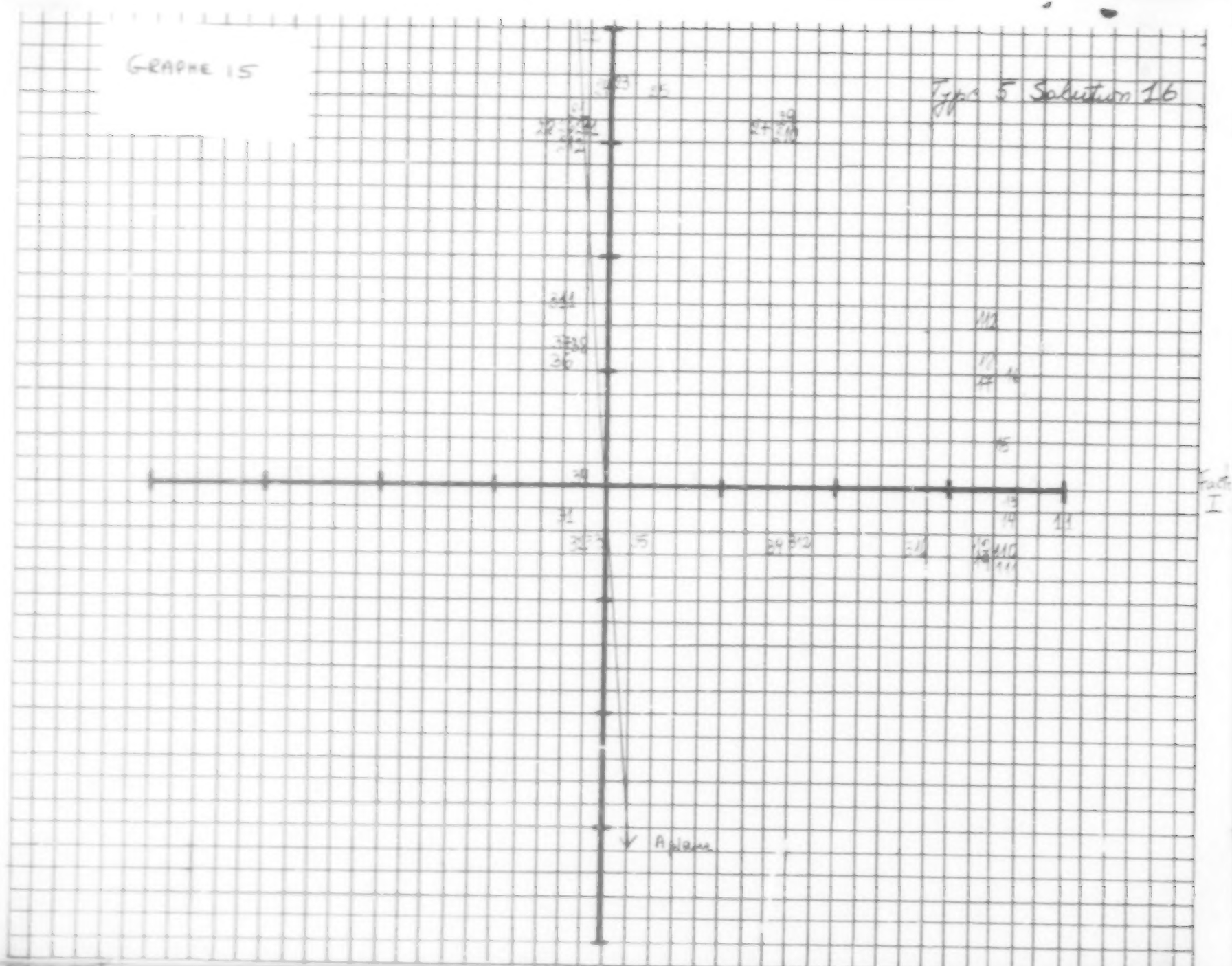
Le juge 1 mentionne que la solution A.F.P(MAXROW)-promax est une bonne solution dans ce cas-ci.

Le juge 2 indique que les solutions suivantes sont "légèrement meilleures" que les autres. Il s'agit des neuf solutions produites par la méthode d'extraction A.C.P.

Le juge 3 spécifie que sa solution privilégiée est A.C.P-varimax.

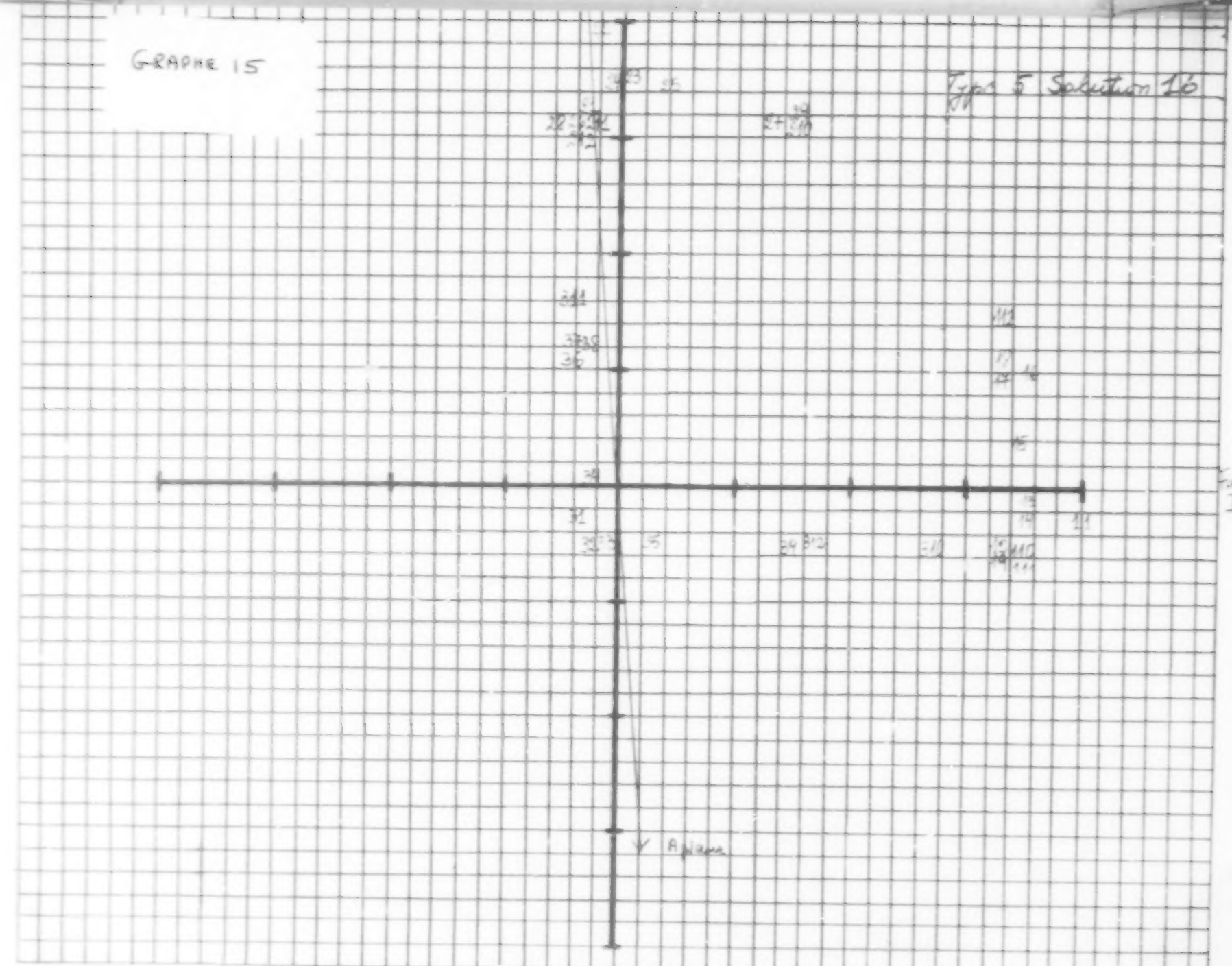
GRAPHE 15

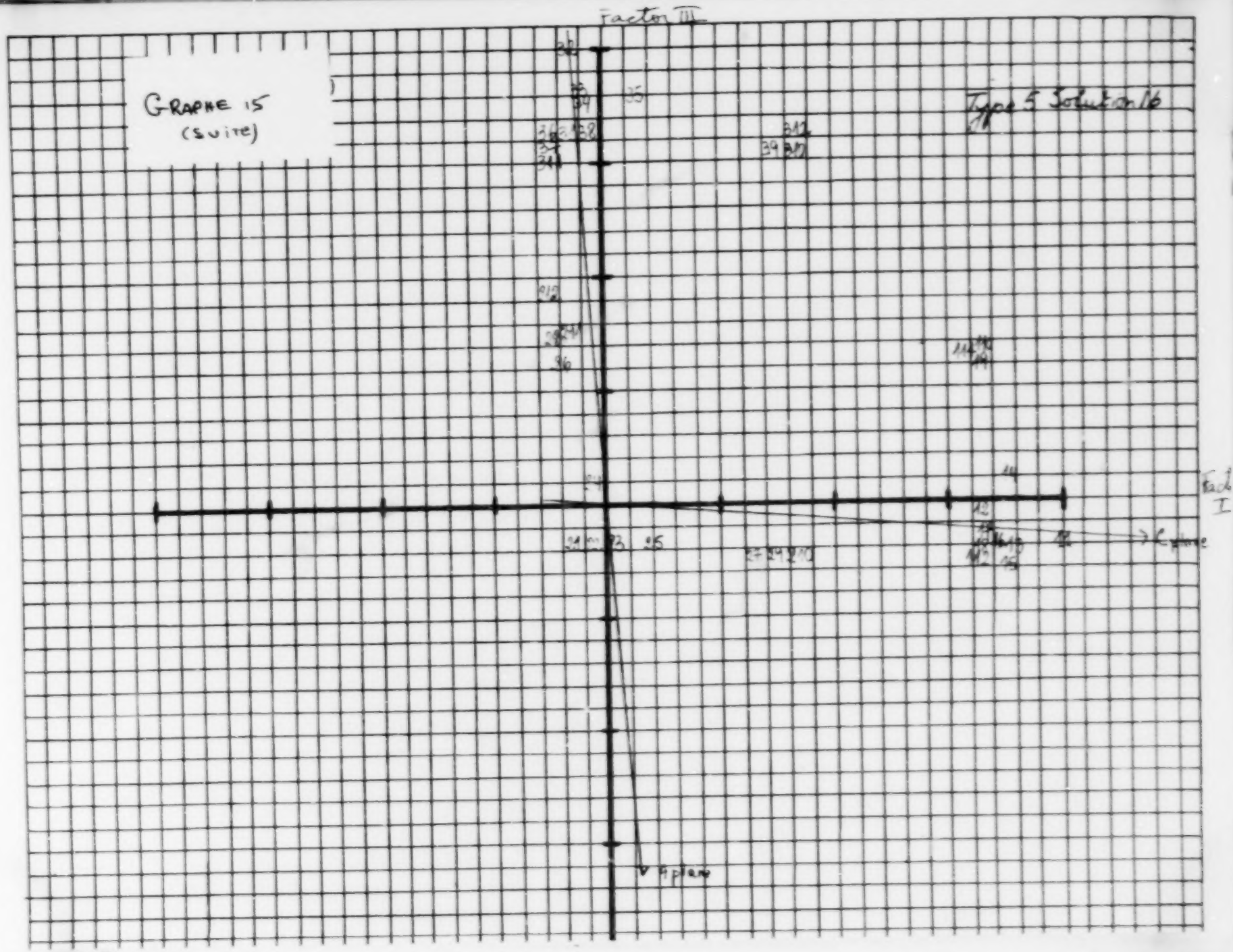
Type 5 Solution 16



GRAPHE 15

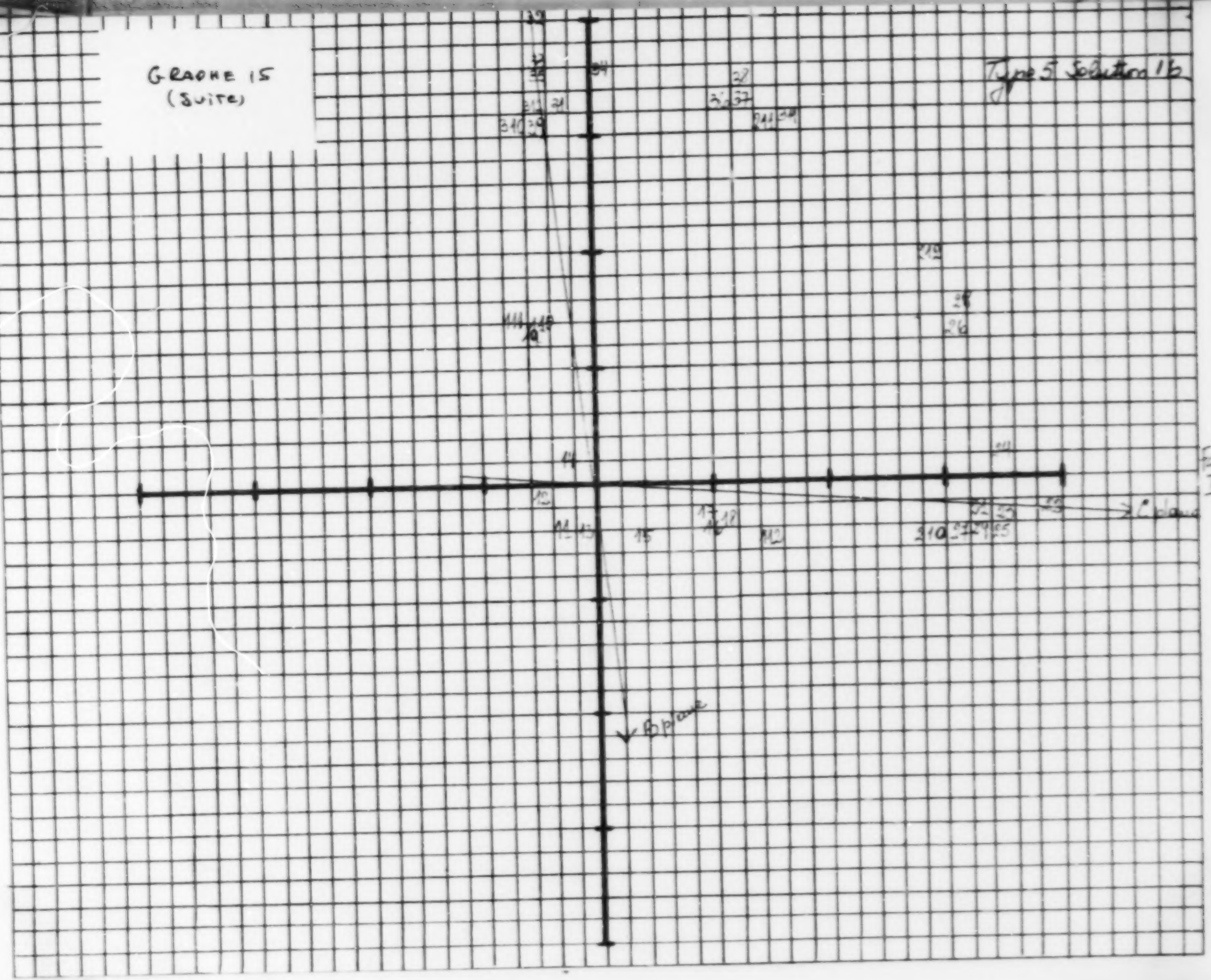
Type 5 Solution 16




Facteur III
GRAPHE 15
(suite)
Type 5 Solution 16
Facteur I
Facteur
Facteur

GRAPHE 15
(Suite)

Type 5 Solution 1b



6.5 Sommaire des résultats

En ce qui concerne le type I, on rencontre une grande hétérogénéité d'opinions quant aux solutions à privilégier (s'il devait y en avoir une). Cependant il serait bon de se rappeler la remarque d'un juge concernant la similarité des solutions. Si l'on excepte les solutions trouvées aberrantes selon le critère 1, soit les cinq orthoblique, les cinq maxplane et malifa-promax, on voit que les solutions proposées par le juge 3 ne sont pas si différentes des quatre premières du juge 2.

Dans le cas du type III, on constate ici aussi, du moins à première vue, une très grande diversité dans le choix des méthodes. Cependant, tel qu'indiqué par un des juges et tel que confirmé par le critère 1, il n'y a pas de solution aberrante pour ce type : elles sont de fait assez semblables. Les choix des juges semblent donc moins hétérogènes que l'on aurait pu le penser suite à une première analyse.

Pour le type V, le pattern des réponses données par les juges est si diversifié que nous n'avons tenté aucune synthèse des résultats à ce moment-ci.

En ce qui a trait au type VII, les opinions sont plus claires. Les choix des juges 2 et 3 sont concordantes. Par ailleurs comme le critère 1 spécifie qu'il n'y a pas de solution aberrante pour ce type, on ne peut conclure que le choix du juge 1 soit si distinct du choix des autres juges.

En résumé, il aurait été bien téméraire de croire que les réponses données par les trois juges seraient tout à fait concordantes et que nous pourrions arriver à un meilleur choix pour chaque type. Comme on

l'a constaté lors de l'analyse du critère 1 et tel qu'indiqué par un des juges, plusieurs solutions sont fort semblables. Ainsi le choix sur une base empirique d'une seule meilleure solution est utopique et peut-être même pas très utile.

La distinction entre les solutions jugées acceptables et celles jugées inacceptables ou encore le choix d'un ensemble de solutions semblables jugées (majoritairement) plus acceptables que d'autres semblent donc des orientations plus prometteuses et plus réalistes.

C'est ce qui sera fait au cours de la prochaine section.

II.7 Discussion et synthèse des résultats

La problématique de l'indétermination des solutions nous a amené à comparer quarante-cinq combinaisons de méthodes d'extraction et de rotation appliquées à huit types de matrices de corrélations. Nous voulions par là identifier la ou les solutions (ou combinaisons de méthodes) qui nous paraissent les plus appropriées pour chacun des types de matrice. Quatre critères d'ordre différent ont été utilisés à cette fin.

Pourtant, aucun consensus évident n'apparaît au terme de cette étude. Aucune solution ou même aucun groupe homogène de solutions ne ressort clairement comme la ou le meilleur(e) pour un type donné. Les résultats, tels qu'exposés au cours des quatre dernières sections sont très diversifiés. Toutefois, il est possible d'extraire de ces résultats un certain nombre de lignes de force ; et ce surtout si nous nous contentons d'identifier non pas la ou les meilleure(s) solution(s) sans aucune équivoque mais plus modestement les solutions qui nous paraissent plus acceptables que d'autres. C'est-à-dire que nous procéderons par élimination, en rejetant d'abord les solutions qui nous apparaissent les moins acceptables.

Voici comment seront déterminées ces solutions jugées inacceptables. Nous les définissons comme étant celles qui ne satisfont pas l'un ou l'autre de nos trois premiers critères. Cela signifie

- pour le critère 1 : solution aberrante
- pour le critère 2 : solution luxueuse (en temps)
- pour le critère 3 : solution sans invariance factorielle.

Nous recouperons ensuite ces informations avec les jugements des spécialistes (critère 4) dans le cas des types I, III, V et VII.

Le tableau 31 indique pour chaque type les solutions trouvées inacceptables selon la définition donnée plus haut : lorsque l'on voit un chiffre comme 1, 2 ou 3 à l'intersection d'une ligne et d'une colonne, cela signifie que la solution (ligne) de ce type (colonne) ne respecte pas l'un ou l'autre des critères 1, 2 ou 3.

Le type I révèle d'après ce tableau trente-quatre solutions qui ont été retenues comme acceptables. Par ailleurs comme les quatre premières solutions du juge 2 ainsi que les solutions du juge 3 se trouvent parmi ces trente-quatre solutions, nous les proposons non pas comme les meilleures mais plutôt comme un groupe de solutions acceptables.

Ainsi pour le type I, les solutions :

A.C.P. [- orthomax ($\gamma = 0$, $\gamma = 0{,}25$, $\gamma = 0{,}50$, $\gamma = 0{,}75$, $\gamma = 1$)
- biquartimin

A.F.P(MAXROW)-orthomax ($\gamma = 0$, $\gamma = 0{,}25$, $\gamma = 0{,}50$, $\gamma = 0{,}75$, $\gamma = 1$)

A.F.P(SMC)-promax

Alpha-promax

Malifa-biquartimin.

sont proposées comme les plus acceptables.

Puisque nous n'avons pas d'informations de spécialistes pour le type II, nous proposerons comme plus acceptables les solutions qui n'ont pas été jugées inacceptables par l'un ou l'autre des trois premiers critères à savoir :

A.C.P. [- orthomax ($\gamma = 0$, $\gamma = 0{,}25$, $\gamma = 0{,}50$)
- maxplane

TABLEAU 31 : Solutions jugées inacceptables[1] pour chacun des types

SOLUTION \ TYPE	I	II	III	IV	V	VI	VII	VIII
1								
2								
3								
4		3						
5		3						
6		3			1	1		
7	1	1 3			1	1		
8		3						
9	1 3							
10								
11								
12								
13		3						
14		3						
15		3			1	1		
16	1	1 3			1	1		
17		3						
18	1 3					3		
19								
20								
21								
22		3						
23		3						
24		3			1	1		
25	1	1 3			1	1		
26		3						
27	1 3					3		

1 Une solution sera jugée inacceptable si elle ne satisfait pas l'un ou l'autre des trois premiers critères déjà définis ; dans lequel cas le numéro du critère est indiqué à l'intersection de la ligne (solution) et de la colonne (type) correspondantes. Par exemple le "3" qui se situe à l'intersection de la ligne "4" et de la colonne "II" signifie que la solution 4 du type II ne respecte pas le critère 3.

TABLEAU 31 : (suite)

SOLUTION \ TYPE	I	II	III	IV	V	VI	VII	VIII
28					2	2	2	2
29					2	2	2	2
30					2	2	2	2
31		3			2	2	2	2
32		3			2	2	2	2
33		3			1 2	1 2	2	2
34	1	1 3			1 2	1 2	2	2
35		3			2	2	2	2
36	1	3			2	2 3	2	2
37		1			2 3	2 3	2	2
38		1			2 3	2 3	2	2
39		1			2 3	2 3	2	2
40		1			2 3	2 3	2	2
41		1			2 3	2 3	2	2
42	1	1 3			2 3	1 2 3	2	2
43	1 3	1 3			1 2 3	1 2 3	2	2
44		1			2 3	2 3	2	2
45	1	1 3		3	1 2 3	2 3	2	2

A.F.P(MAXROW) [- orthomax ($\gamma = 0$, $\gamma = 0{,}25$, $\gamma = 0{,}50$)
- maxplane

A.F.P(SMC) [- orthomax ($\gamma = 0$, $\gamma = 0{,}25$, $\gamma = 0{,}50$)
- maxplane

alpha-orthomax ($\gamma = 0$, $\gamma = 0{,}25$, $\gamma = 0{,}50$)

En procédant de la sorte pour les autres cas, on obtient comme solutions plus acceptables pour le type III.

A.C.P. [- orthomax ($\gamma = 0$, $\gamma = 0{,}25$, $\gamma = 0{,}50$, $\gamma = 0{,}75$, $\gamma = 1$)
- maxplane

A.F.P(MAXROW) [- promax
- biquartimin
- maxplane

A.F.P(SMC) [- biquartimin
- maxplane

Alpha-maxplane

Malifa-maxplane

Pour le type IV, toutes les solutions sauf malifa-maxplane.

Pour le type V,

A.C.P.-orthomax ($\gamma = 0$, $\gamma = 0{,}25$, $\gamma = 0{,}50$, $\gamma = 0{,}75$, $\gamma = 1$)
A.F.P(MAXROW)-orthomax ($\gamma = 0$, $\gamma = 0{,}25$, $\gamma = 0{,}50$, $\gamma = 0{,}75$, $\gamma = 1$)
A.F.P(SMC)-orthomax ($\gamma = 0$, $\gamma = 0{,}25$, $\gamma = 0{,}50$, $\gamma = 0{,}75$, $\gamma = 1$)

Pour le type VI,

A.C.P.
- orthomax ($\gamma = 0$, $\gamma = 0{,}25$, $\gamma = 0{,}50$, $\gamma = 0{,}75$, $\gamma = 1$)
- biquartimin
- maxplane

A.F.P.
- (MAXROW)-orthomax ($\gamma = 0$, $\gamma = 0{,}25$, $\gamma = 0{,}50$, $\gamma = 0{,}75$, $\gamma = 1$)
- biquartimin
- promax (modifiée comme au graphe 15)

A.F.P.(SMC)
- orthomax ($\gamma = 0$, $\gamma = 0{,}25$, $\gamma = 0{,}50$, $\gamma = 0{,}75$, $\gamma = 1$)
- biquartimin

Pour le type VII,

toutes les A.C.P et A.F.P(MAXROW)- promax.

Pour le type VIII,

toutes les solutions A.C.P., A.F.P(MAXROW), A.F.P.(SMC).

En plus de proposer les solutions qui nous paraissent comme les plus acceptables pour un type donné de matrice de corrélations, nous allons mettre en évidence certains résultats plus généraux qui semblent se dégager des analyses précédentes. Une liste de recommandations suivra cette discussion.

En scrutant les lignes du tableau 31 on remarque que très peu de solutions (lignes) sont complètement exemptes de chiffres. Ces solutions n'ont donc jamais été trouvées inacceptables, que ce soit pour l'un ou l'autre type. En fait, seulement neuf solutions ont cette qualité. Il s'agit de A.C.P.-orthomax ($\gamma = 0$, $\gamma = 0{,}25$, $\gamma = 0{,}50$), A.F.P(MAXROW)-orthomax ($\gamma = 0$, $\gamma = 0{,}25$, $\gamma = 0{,}50$) et A.F.P(SMC)-orthomax ($\gamma = 0$, $\gamma = 0{,}25$, $\gamma = 0{,}50$). Cela signifie que ces solutions sont à la fois semblables à leur solution médiane respective, peu coûteuse, et invariante psychométriquement ; et ce quel que soit le type de données utilisé.

Par contre, certaines solutions sont jugées inacceptables pour la plupart des types de données analysées. La pire performance est celle de malifa-maxplane qui fut jugée inacceptable, 7 fois sur 8. Parmi les solutions jugées inacceptables 6 fois sur 8, on note alpha-orthoblique, alpha-maxplane, malifa-promax et malifa-orthoblique. Il faut dire qu'en général les méthodes d'extraction alpha et surtout malifa donnent une bonne part de solutions inacceptables. À la lumière de nos résultats, on peut même affirmer qu'il serait bien téméraire de choisir de telles méthodes d'extraction, surtout pour des matrices de corrélations ayant plus de trente variables. Exception faite de quelques rares utilisations spécifiques qui requièrent des méthodes d'extraction de cette nature (Thorndike [1978, p. 324] et Massey [1982]). D'autant plus que la méthode malifa suscite encore de nos jours un certain nombre d'interrogations de tous ordres chez les chercheurs (Olsson [1979], Jackson et Chan [1980], Bentler [1982]).

Les résultats de rotation obliques sont plus souvent jugées inacceptables que les méthodes orthogonales, sauf pour biquartimin qui se comporte plutôt comme les orthogonales à ce chapitre. Les bons résultats obtenus avec cette méthode oblique ne sont pas les premiers (Harman [1976, p. 313]), même si certains auteurs ne la recommandent pas (Gorsuch [1974, p. 212]).

Remarquons par ailleurs au tableau 31 que le pattern des solutions inacceptables est identique, à une seule exception près, pour les méthodes d'extraction A.C.P., A.F.P(MAXROW) et A.F.P(SMC).

Nous mettons ici un terme à cette discussion des résultats en suggérant une liste de recommandations inspirées par nos analyses.

II.8 Recommandations

- Le choix du progiciel (SPSS, BMDP, SAS, OSIRIS, IMSL...) ne doit pas se faire à l'aveuglette. Les questions de temps d'exécution informatique (reliées au coût d'une analyse) sont tributaires de ce choix, comme on l'a déjà constaté.

- D'après nos observations, il n'y a pas beaucoup de risques à utiliser comme méthode d'extraction l'A.C.P. et comme méthode de rotation une des orthomax pour n'importe quel type de données similaires à celles que nous avions. Les solutions obtenues à partir de la combinaison de ces méthodes d'extraction et de rotation ont souvent été jugées parmi les plus acceptables pour les huit types de données.

- On devrait avoir des raisons bien précises pour employer une des méthodes d'extraction alpha ou malifa, surtout lorsque les matrices de corrélations comprennent plus de 30 variables. Les solutions provenant de ces méthodes ont souvent été trouvées inacceptables selon nos critères.

- Nous ne pouvons recommander chaudement les méthodes de rotation oblique avec les données que nous avons en mains et les critères que nous avons employés. En général les solutions provenant de ces méthodes sont jugées plus souvent inacceptables que les méthodes orthogonales, si on fait exception de la méthode oblique biquartimin qui se comporte comme les orthogonales.

- Lorsque le nombre de variables d'une matrice dépasse 30, il est important de choisir judicieusement la bonne méthode d'extraction surtout si des questions de coût (ou de temps) sont en jeu.

- Autant que possible, on devrait viser à choisir des matrices à analyser comme celles des types III, IV, VII, VIII, dites de corrélations faibles. Toutes les solutions de ces types, sauf une, respectent parfaitement les critères 1 et 3.

- Si le respect du critère d'invariance factorielle nous importe beaucoup, on devrait tenter d'obtenir le même nombre de variables par groupe (ou facteur) et donc utiliser des matrices de type I, III, V ou VII.

CONCLUSION

Au terme de cet exposé qui a brièvement présenté en première partie quelques méthodes d'analyse factorielle et qui a relaté en seconde partie une étude comparative de plusieurs de ces méthodes, qu'il nous soit permis de poser un regard quelque peu perplexe quant aux résultats obtenus.

Non que le bilan de cette étude soit vraiment négatif. Certes nous n'avons pas trouvé la meilleure solution pour chacun des types de matrice de corrélations étudiés tel que nous l'escomptions peut-être un peu naïvement au début de cette étude comparative. Mais est-ce bien là le véritable enjeu ? Nous sommes plutôt portés à croire aujourd'hui comme Kim et Mueller [1978b] qu'il n'existe pas une solution définitive (ou meilleure solution) pour la majorité des types de matrices. Ce qui ne veut pas dire que toutes les solutions sont bonnes ou acceptables universellement : non, notre étude comparative le montre bien.

Cependant cet ouvrage montre tout aussi bien qu'il pourrait exister un compromis entre "utiliser n'importe quelle combinaison de méthodes d'extraction et de rotation" ou pire "utiliser toujours la même (ex. A.C.P.-varimax)" et "choisir la meilleure solution pour chaque type de données". Autrement dit, même s'il est peu probable que nous trouvions une solution définitive pour chaque type de données, il serait inconséquent et irresponsable d'aller à l'opposé et de ne jamais utiliser de jugement critique pour effectuer un choix de méthode. Il nous apparaît par exemple que l'utilisation des méthodes d'un seul progiciel peut restreindre considérablement ce choix.

Pour conclure nous sommes d'avis que tout bon "utilisateur modéré" devrait pouvoir discriminer entre les méthodes acceptables et celles qui ne le sont pas compte tenu des données dont il dispose. Il est à espérer que les résultats obtenus au cours de notre étude comparative sauront l'aider à y parvenir.

Car si nous n'avons pas obtenu toutes les réponses escomptées au tout début de cette étude, du moins avons-nous trouvé, nous semble-t-il, la bonne question à se poser quant au choix d'une méthode : "Est-ce que cette méthode paraît acceptable ?" semble en effet plus réaliste, plus modeste et plus légitime que "Quelle est la meilleure solution ?".

BIBLIOGRAPHIE

AJAR, D., [1982] "Le problème de la détermination du nombre de facteurs en analyse factorielle", Revue des sciences de l'éducation, Vol. VIII, No 1.

ALLEN, L., [1978] "Factor Analysis of the Ward Atmosphere Scale", Journal of Consulting and Clinical Psychology, Vol. 46, No 1.

ANDERSON, T.W., DAS GUPTA, S.D., STYAN, G.P.H., [1977] A Bibliography of Multivariate Statistical Analysis", Krïegrer Publishing Company, Huntington, N.Y.

BENTLER, P.M., [1982] "Confirmatory Factor Analysis Via Non iterative Estimation : a Fast Inexpensive Method", Journal of Marketing Research, Novembre

BERTRAND, R., [1977] "Solutions universelles en analyse factorielle", INRS-Éducation, document C-052.

BERTRAND, R., DUFOUR, N., [1981] "Factorial Invariance of Orthomax Rotations : a Theoritical-Experimental Study", Interdisciplinaria, Vol. 2, No 2.

BROWNE, M.W., [1968] "A Comparison of Factor Analytic Techniques" Psychometrika, Vol. 33, No 3.

BUTLER, J.J., [1969] "Simple Structure Reconsidered : Distinguishability and Factorial Invariance in Factor Analysis", Multivariate Behavioral Research, Vol. 4.

CAILLIEZ, F., PAGES, J.P., [1976] Introduction à l'analyse des données, SMASH, Paris.

CARROLL, J.B., [1957] "Biquartimin Criterion for Rotation to Oblique Simple Structure in Factor Analysis", Science, Vol. 126.

CARROLL, J.B., [1960] "IBM 704 Program for Generalized Analytic Rotation Solution in Factor Analysis", Harvard University, Cambridge, Mass.

CATTELL, R.B., KHANNA, D.K., [1977] Principles and Procedures for Unique Rotation in Factor Analysis, in Statistical Methods for Digital Computers, Vol. III, (Enslein, Ralston & Wilf, eds), Wiley-Interscience, N.Y.

CATTELL, R.B., MUERLE, J.L., [1960] "The Maxplane Program for Factor Rotation to Oblique Simple Structure", Educational and Psychological Measurement, Vol. 20.

COCHÉ, E., MEEHAN, J., [1979] "Factor and Cluster Analysis with the Rosenzweig Picture Frustration Study" Journal of Personality Assessment, Vol. 43, No 1.

CONGER, A.J., STALLARD, E., [1976] Equivalencies Among Canonical Factor Analysis, Canonical Reliability Analysis and Principal Components : Implications for Data Reduction of Fallible Measures", Educational and Psychological Measurement, Vol. 36.

CURETON, E.E., D'AGOSTINO, R.B., [1983] Factor Analysis : An Applied Approach, Lawrence Erlbaum Associates Publishers, Hillsdale, N.J.

DIELMAN, T.E., CATTELL, R.B., WAGNER, A., [1972] "Evidence on the Simple Structure and Factor in Variance Achieved by Five Rotational Methods on Four Types of Data", Multivariate Behavioral Research, Vol. 7.

DORANS, N.J., [1977] "Factorial Validity by Tautology : Methodological Comments on Stewart and Stewart" Multivariate Experimental Clinical Research, Vol. 3, No 2.

EBER, H.W., [1966] "Toward Oblique Simple Structure : Maxplane" Multivariate Behavioral Research, Vol. 1.

EYMAN, R.K., DINGMAN, H.F., MEYERS, C.E., [1962] "Comparison of Some Computer Techniques for Factor Analytic Rotation", Educational and Psychological Measurement, Vol. 22.

GORSUCH, R.L., [1970] "A Comparison of Biquartimin, Maxplane, Promax and Varimax", Educational and Psychological Measurement, Vol. 30.

GORSUCH, R.L., [1974] "Factor Analysis", Saunders, Philadelphie.

HARMAN, H.H., [1976] Modern Factor Analysis, The University of Chicago Press, Chicago.

HARRIS, M.L., HARRIS, C.W., [1971] "A factor Analytic Interpretation Strategy", Educational and Psychological Measurement, Vol. 31, No 3.

HARRIS, C.W., KAISER, H.F., [1964] "Oblique Factor Analytic Solutions By Orthogonal Transformations", Psychometrika, Vol. 29, No 4.

HENDRICKSON, D.E., WHITE, P.O., [1964] "Promax : A Quick Method for Rotation to Oblique Simple Structure", British Journal of Mathematical and Statistical Psychology, Vol. 17.

HOFMANN, R.J., [1978] "The Orthotran Solution", Multivariate Behavioral Research, Vol. 13.

HOLZINGER, K.J., [1937] "The Bi-factor Method", Psychometrika, Vol. 2.

HORN, J.L., [1967] "On Subjectivity in Factor Analysis" Educational and Psychological Measurement, Vol. 27.

HOTELLING, H., [1933] "Analysis of a Complex of Statistical Variables into Principal Components", Journal of Educational Psychology, Vol. 24.

HOWARTH, E., [1977] "Mood Variation Patterns Based on Factor Analysis", Psychological Reports, Vol. 41.

HUMPHREYS, L.G., ILGEN, D.R., [1969] "Note on a Criterion for the Number of Common Factors", Educational and Psychological Measurement, Vol. 29.

HUMPHREYS, L.G., MONTANELLI, R.G., [1975] "An Investigation of the Parallel Analysis Criterion for Determining the Number of Common Factors", Multivariate Behavioral Research, Vol. 10.

HURLEY, J.R., CATTELL, R.B., [1962] "The Procrustes Program, Producing Direct Rotation to Test a Hypothesized Factor Structure, Behavioral Science, Vol. 7.

JACKSON, D.J., BORGATTA, E.F., GOLDSMITH, H.F., [1979] "Data Analysis in Factorial Ecology" Sociological Methods and Research, Vol. 7, No 3.

JACKSON, D.N., CHAN, D.W., [1980] "Maximum-likelihood Estimation in Common Factor Analysis ; A Cautionary Note", Psychological Bultin, Vol. 88, No. 2.

JÖRESKOG, K.G., [1969] "A General Approach to Confirmatory Maximum Likelihood Factor Analysis", Psychometrika, Vol. 34.

KAISER, H.F., [1958] "The Varimax Criterion for Analytic Rotation in Factor Analysis", Psychometrika, Vol. 23.

KIM, J.O., MUELLER, C.W., [1978a] Introduction to Factor Analysis : What it is and How to Do it, Sage Publications, Beverly Hills, Californie.

KIM, J.O., MUELLER, C.W., [1978b] Factor Analysis : Statistical Methods and Pratical Issues, Sage Publications, Beverly Hills, Californie.

LAUX, F.D., TUCKER, R.K., [1979] "The effects of Differentially Computed Factor Scores on Statistical Decisions", The Psychological Record, Vol. 29.

LAWLEY, D.N., MAXWELL, A.E., [1971] Factor Analysis as a Statistical Method, Butterworth, London.

LEE, H.B., COMREY, A.L., [1978] "An Empirical Comparison of two Minimum Residual Factor Extraction Methods", Multivariate Behavioral Research, Vol. 13.

MANNERS, G.E., BRUSH, D.H., [1979] A Simulation of Factor Analytic Reliability Varying Sample Sizes and Number of Variables, Psychological Reports, Vol. 45.

MASSEY, A.J., [1982] "Analysing the Dimensionality of Achievement Test Data : an Illustration Example Using Factor Analysis", CORE, Vol. 6, No 2.

MC DONALD, R.P., MULAIK, S.A., [1979] "Determinacy of Common Factors : A Non Technical Review", Psychological Bulletin, Vol. 86, No 2.

MORRIS, J.D., [1979] "A Comparison of Regression Prediction Accuracy on Several Types of Factor Scores", American Educational Research Journal, Vol. 16, No 1.

MULAIK, S.A., [1972] "The Foundations of Factor Analysis, McGraw-Hill, New-York.

NEUHAUS, J.O., WRIGLEY,C., [1954] "The Quartimax Method : An Analytical Approach to Orthogonal Simple Structure", British Journal of Statistical Psychology, Vol. 7.

OLSSON, U., [1979] "On the Robustness of Factor Analysis Against Crude Classification of the Observations", Multivariate Behavioral Research, Vol. 14.

REUCHLIN, M., [1964] Méthodes d'analyse factorielle à l'usage des psychologues, Presses Universitaires de France, Paris.

REVELLE, W., ROCKLIN, T., [1979] "Very Simple Structure : An Alternative Procedure for Estimating the Optimal Number of Interpretable Factors", Multivariate Behavioral Research, Vol. 14.

RHÉAUME, D., [1982] "La diversité des intérêts", INRS-Éducation, document R-149.

RIMOLDI, H.J.A., BÉGIN, Y. et alii [1982] "Students' Psychological Characteristics and Problem Solving : a Cross-Cultural Study", Interdisciplinaria, Monographie No 1.

RUMMEL, R.J., [1970] Applied Factor Analysis, Norwestern University Press, Evanston.

SPEARMAN, C., [1904] "General Intelligence Objectiverly Determined and Measured", American Journal of Psychology, Vol. 15.

STEIGER, J.H., [1979a] "The Relationship Between External Variables and Common Factors", Psychometrika, Vol. 44, No 1.

STEIGER, J.H., [1979b] "Factor Indeterminacy in the 1930's and in the 1970's : Some Interesting Parallels", Psychometrika, Vol. 44, No 1.

STEIGER, J.H., [1980] "On the Empirical Equivalence of Regression Components and Common factors", British Journal of Mathematical and Statistical Psychology, Vol. 33.

TATSUOKA, M.M., [1971] Multivariate Analysis, John Wiley and Sons, N.Y.

THORNDIKE, R.M., [1978] Correlational Procedures for Research, Gardner Press, N.Y.

THURSTONE, L.L., [1931] "Multiple Factor Analysis", Psychological Review, Vol. 38.

THURSTONE, L.L., [1947] Multiple Factor Analysis, University of Chicago Press, Chicago.

TUCKER, L.R., KOOPMAN, R.F., LINN, R.L., [1969] "Evaluation of Factor Analytic Research Procedures by Means of Standard Correlation Matrices", Psychometrika, Vol. 34.

VELICER, W.F., [1977] "An Empirical Comparison of the Similarity of Principal Component, Image, and Factor Patterns", Multivariate Behavioral Research, Vol. 12.

WARBURTON, F.W., [1963] "Analytic Methods of Factor Rotation", British Journal of Statistical Psychology, Vol. 16.

ANNEXE I

Calcul des saturations du pattern factoriel et des coefficients des scores factoriels pour les données du tableau 1 selon le modèle de l'analyse en composantes principales.

Nous partons de la matrice des corrélations

$$R = \begin{pmatrix} 1 & 0,83397 \\ 0,83397 & 1 \end{pmatrix}.$$

La première étape du calcul consiste à trouver les valeurs propres (eigenvalues) λ_1 et λ_2 de cette matrice. On les obtient en solutionnant l'équation

$$\begin{vmatrix} 1 - \lambda & 0,83397 \\ 0,83397 & 1 - \lambda \end{vmatrix} = 0$$

i.e. $\lambda^2 - 2\lambda + 0,30449 = 0$

Les racines de cette équation sont $\frac{2 \pm 1,66794}{2} = 1 \pm 0,83397$

Ainsi $\lambda_1 = 1,83397$ $\lambda_2 = 0,16603$

Ce qui donne la ligne "variance du facteur" du pattern factoriel au tableau 2.

Les saturations a_{ij} du pattern factoriel sont obtenus de la façon suivante.

Il s'agit de résoudre les équations matricielles

$$R \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix} = \lambda_1 \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix}$$

et $$R \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix} = \lambda_2 \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix}$$

i.e. $$\begin{pmatrix} 1 & 0,83397 \\ 0,83397 & 1 \end{pmatrix} \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix} = \begin{pmatrix} 1,83397\ a_{11} \\ 1,83397\ a_{21} \end{pmatrix}$$

et $$\begin{pmatrix} 1 & 0,83397 \\ 0,83397 & 1 \end{pmatrix} \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix} = \begin{pmatrix} 0,16603\ a_{12} \\ 0,16603\ a_{22} \end{pmatrix}$$

Ce qui donne

$$a_{11} + 0,83397\ a_{21} = 1,83397\ a_{11}$$

et $$a_{12} + 0,83397\ a_{22} = 0,16603\ a_{12}$$

i.e. $$a_{11} - a_{21} = 0 \quad (1)$$

et $$a_{12} + a_{22} = 0 \quad (2)$$

Or comme $\lambda_1 = a_{11}^2 + a_{21}^2$ ③

et $\lambda_2 = a_{12}^2 + a_{22}^2$ ④

on aura, en remplaçant ① dans ③ et ② dans ④ ,

$$2a_{11}^2 = \lambda_1 = 1,83397$$

$$2a_{12}^2 = \lambda_2 = 0,16603$$

i.e. $a_{11} = 0,95759$, d'où $a_{21} = 0,95759$ d'après ①
et $a_{12} = -0,28812$, d'où $a_{22} = + 0,28812$ d'après ② .

(on choisit a_{12} négatif pour que l'axe Z_G (graphe 3) soit orienté vers le haut).

Les coefficients des scores factoriels sont obtenus en divisant chacune des saturations correspondante du pattern factoriel par la variance du facteur (i.e. la valeur propre)

Ainsi $\frac{0,95759}{1,83397} = 0,52214$ et $\frac{0,28812}{0,16603} = 1,73535$

Les justifications théoriques de ces calculs se trouvent dans Harman [1976, chap. 8].

ANNEXE II

Le traitement informatique

- L'extraction de facteurs

Deux progiciels, SPSS, et BMDP, ont dû être employés pour produire les facteurs. En effet aucun progiciel à lui seul ne couvrait toutes les méthodes à l'étude.

Le progiciel SPSS a été utilisé pour les méthodes d'extraction, l'analyse en composantes principales et la méthode alpha. Le progiciel BMDP a servi pour les méthodes d'extraction des facteurs principaux (SMC et MAXROW) et la méthode du maximum de vraisemblance.

Les paramètres des progiciels comme les critères de convergence, le nombre de facteurs retenus et la tolérance à l'inversion matricielle sont dans la plupart des cas les standards préétablis par les programmes. Les déviations à ces normes ont été peu fréquentes sauf pour la méthode d'extraction de maximum de vraisemblance appliquée aux matrices de types I, II, V et VI, où la tolérance à l'inversion matricielle a été souvent relâchée pour permettre l'extraction de facteurs.

- La rotation des facteurs

Le progiciel IMSL a été utilisé pour les cinq méthodes de rotation orthogonale ainsi pour les méthodes de rotation oblique promax et

orthoblique. Pour les méthodes biquartimin et maxplane, les programmes utilisés ont été construits en se basant sur ceux présentés par Guertin et Bailey[1] [1970].

Cette démarche a été adoptée de préférence au développement complet de nos propres programmes compte tenu de la grande difficulté de valider de façon satisfaisante de tels programmes. De plus, la plus grande accessibilité des programmes que nous avons choisis rendra plus aisée la reprise de notre étude par d'autres chercheurs.

Nous décrirons maintenant les paramètres pertinents à chacune des méthodes de rotation.

Notons que la normalisation de Kaiser a été utilisée dans tous les cas.

- Rotations orthogonales : le nombre d'itération maximum est fixé à 30 tandis que les deux critères de convergence FPS et delta sont respectivement 0,00001 et 0,00001.

- Promax : la rotation orthogonale préliminaire est varimax. La puissance à laquelle on élève les facteurs a été fixée à quatre.

- Orthoblique : la rotation orthogonale préliminaire est varimax. La puissance à laquelle on élève la matrice diagonale des racines carrées des valeurs propres a été fixée à zéro pour conserver la méthode originale proposée par Harris et Kaiser.

---

1 GUERTIN, W.H., BAILEY, J., [1970] Introduction to Modern Factor Analysis, Edwards Brothers Inc. An Arbor.

. Biquartimin : le critère de convergence est fixé à 0,00001 et le cosinus maximum de l'angle entre deux facteurs se situe entre 0,94 et 0,97.

. Maxplane : la largeur de l'hyperplan a été de 0,10, la pondération par la saturation du co-facteur ("faith") de 2,00, le critère de convergence était 0,001 et enfin l'angle minimum entre les facteurs de 45°.

D'autre part, l'angle de rapprochement des facteurs dans le processus itératif était de 2° pour les types I, II, III, IV et de 6° pour les types V, VI, VII, VIII.

Université du Québec
Institut national de la recherche scientifique
INRS-Éducation
2383, chemin Sainte-Foy
Sainte-Foy (Québec)
G1V 1T1
Téléphone: (418) 657-2560,
poste 2890